

Stato dell'arte per le tecnologie del Non Intrusive Load Monitoring e del Clustering delle utenze

Documento di indirizzo per le attività del progetto SMILE
nell'ambito dei servizi di consulenza del CRS4 per Energia Mediterranea S.r.l.

Luca Massidda, CRS4 - 20 Settembre 2017

INDICE

I	Introduzione	1
II	Tecniche di pattern matching energetico ed individuazione dei carichi	1
II-A	Metodologie di modellazione e trattamento	3
II-B	Benchmarking	4
II-C	Librerie	4
III	Tecniche di profilazione e clustering degli utilizzatori energetici	4
III-A	Pre-clustering	4
III-B	Data reduction	4
III-C	Clustering	5
III-D	Classificazione	5
IV	Dataset pubblici	5
	Riferimenti bibliografici	5

I. INTRODUZIONE

Buona parte dei consumi elettrici attuali è legata alle utenze domestiche e commerciali, che possono rappresentare più del 50% del totale. Rappresentano quindi anche una interessantissima opportunità per realizzare interventi di razionalizzazione dei consumi, alcuni studi infatti stimano che solo fornendo alle utenze le informazioni continue ed accurate circa i loro consumi si possano realizzare risparmi nel consumo energetico anche del 12%.

Le piccole utenze domestiche e le utenze commerciali rappresentano quindi un mercato interessantissimo e potenzialmente enorme per sistemi di misura dei consumi. Vantaggi ulteriori possono aspettarsi se l'informazione fornita agli utenti non sia limitata al solo dato relativo al consumo ma possa includere una analisi del profilo di consumo, per l'individuazione delle abitudini dell'utente e l'individuazione di eventuali correzioni con un ottica di risparmio; oppure fornendo una analisi dettagliata relativa ai consumi dei singoli apparati elettrici collegati alla rete domestica.

Gli utenti finali non sono generalmente interessati alla migrazione da un sistema analogico ad uno digitale e quindi alla sola conoscenza e disponibilità continua delle informazioni circa i propri consumi, ma richiedono dei meccanismi di

analisi dei dati che possa prefigurare una riduzione dei loro consumi e relativi costi [4].

Allo stato attuale esistono in commercio diverse soluzioni in cui, di solito, le cosiddette smart plugs sono collegate in modo wireless ad un aggregatore di dati. Con smart plug si intende un dispositivo di misura che si inserisce tra una normale presa dell'impianto domestico e l'elettrodomestico di cui si vuole monitorare il consumo, oppure che sostituisce integralmente la presa elettrica standard.

Le smart plug misurano il consumo elettrico in termini di potenza attiva. Tutti i dati sono raccolti da un aggregatore in una rete domestica e sono resi disponibili all'utente, attraverso interfacce grafiche realizzati in apparecchi a se stanti o consultabili tramite browser o specifiche app per i dispositivi mobili.

Questo tipo di misurazione presenta il vantaggio di una apparente facilità di realizzazione, di contro però ha un costo elevato, per la mole di sensori e dispositivi di comunicazione da gestire, ed è comunque invasivo per una utenza già esistente. Per tale motivo la diffusione di questi sistemi è stata finora limitata. Vi è inoltre uno studio che mostra come questo tipo di sistemi non riesca a dare informazioni circa il consumo complessivo di un utenza con il risultato netto di un incremento dei consumi di quasi il 16% [36].

Il monitoraggio dei consumi e la loro catalogazione può anche avvenire in modo non intrusivo attraverso un singolo punto di misura, relativa all'intera utenza, attraverso la analisi del segnale di potenza consumata da tutte le utenze. In questo caso il sistema di misurazione risulta certamente non intrusivo per l'utente, di costo relativo inferiore, ma richiede una sofisticata analisi del dato del consumo da realizzare attraverso algoritmi di intelligenza artificiale.

Gli smart meter stanno incontrando una crescente diffusione in tutto il mondo e, nel contesto delle smart grid, possono offrire l'opportunità agli operatori di avere un maggiore controllo sulla distribuzione energetica, attraverso il monitoraggio dell'uso dei singoli utilizzatori e della produzione da fonte rinnovabile. Con le strategie denominate *demand response* è possibile ridurre i picchi di domanda energetica eliminando i consumi o spostandoli temporalmente [46] [42].

II. TECNICHE DI PATTERN MATCHING ENERGETICO ED INDIVIDUAZIONE DEI CARICHI

La disaggregazione dei carichi, il pattern matching energetico, sono comunemente indicati in letteratura come Non-

Intrusive Load Monitoring (NILM). Si intende con questo il processo di disaggregazione del consumo dei singoli elettrodomestici a partire dal dato relativo al consumo aggregato di utenze domestiche e commerciali.

La nascita di questo filone di ricerca si attribuisce generalmente al lavoro di George Hart, che per primo utilizzò l'analisi dei transienti di potenza attiva e reattiva per individuare l'accensione e lo spegnimento degli elettrodomestici di una casa [26].

Da allora sono stati pubblicati molti lavori di ricerca, che possono essere categorizzati in due gruppi, che si distinguono per le quantità misurate e per la frequenza di campionamento dei segnali utilizzati nell'analisi:

- alta frequenza di campionamento (da 50 Hz oltre il kHz)
- bassa frequenza di campionamento (inferiore a 1 Hz)

Il primo gruppo prevede l'individuazione degli eventi di accensione e spegnimento attraverso l'acquisizione di segnali ad alta frequenza superiore all'hertz, attraverso hardware dedicato. Le tecniche utilizzate vanno dall'individuazione del transiente dovuto ad un interruttore meccanico, analisi del transiente delle armoniche, l'interferenza elettromagnetica generata dai singoli elettrodomestici ed altre [18] [25].

L'approccio seguito consiste in genere nell'individuazione di *firme* nel contenuto armonico dell'alterazione di segnale prodotta da un elettrodomestico sia nel suo funzionamento a regime sia durante le fasi di accensione e spegnimento. L'individuazione di queste firme permette poi l'individuazione degli eventi di accensione e spegnimento a partire dal segnale complessivo sfruttando tecniche di pattern recognition [9] [10].

Questo tipo di metodi oltre ad essere intrusivi e richiedere hardware dedicato non sono compatibili con le caratteristiche comunemente disponibili in uno smart meter.

Recentemente la ricerca sulla disaggregazione dei carichi è focalizzata, invece, sulle metodologie di individuazione dei carichi a partire dal segnale del consumo aggregato a bassa frequenza.

Tipicamente gli smart meter sono in grado di fornire informazioni circa il consumo aggregato di una abitazione o di un locale commerciale con una risoluzione temporale dell'ordine del secondo, si perde in questo modo tutta l'informazione relativa alla forma dell'onda e spesso è disponibile solo la componente attiva della potenza.

I dati utilizzati hanno un periodo di campionamento che va da 1 s ad 1 h, e sono costituiti da quantità ottenibili da smart meter installati nell'utenza in esame, e quindi essenzialmente misurazioni di potenza attiva e reattiva.

Ci si basa in questo caso su letture direttamente ottenibili dagli smart meter, risultando quindi minimamente invasivo per l'utente, ma, a causa della bassa frequenza di misura, con l'incapacità di individuare gli elettrodomestici che hanno un consumo limitato e di breve durata. Tuttavia utenze a consumo elevato e lunga durata, come uno scaldabagno elettrico o una lavatrice, possono essere individuate con una precisione ragionevole anche con frequenze di campionamento di 15min [29] [43] [6] [5].

Le tecniche utilizzabili per la disaggregazione dei carichi in questo contesto devono pertanto essere differenti rispetto a quelle utilizzate nel caso ad alta frequenza.

In generale si possono classificare le tecniche di riconoscimento in due categorie:

- tecniche basate sul riconoscimento degli eventi
- tecniche basate sul riconoscimento dello stato

Gli approcci event based sono focalizzati sulle transizioni legate all'utilizzo degli elettrodomestici, ed utilizzano algoritmi per la change detection per identificare gli eventi. Il change detection algorithm ha il compito di indentificare l'evento di accensione o spegnimento di un elettrodomestico o l'evento per cui questo cambi il proprio stato. Il cambio nella potenza assorbita da una utenza (positiva o negativa) se sufficientemente significativo, rappresentano l'indicazione di un evento. Una sequenza di eventi rappresenta una "firma" per un elettrodomestico, queste firme possono essere riconosciute nella serie temporale dei consumi ed utilizzate per stimare il consumo dei singoli elettrodomestici.

La performance di questi metodi è limitata dal change detection algorithm e dalla soglia per il riconoscimento delle variazioni, dalla rumorosità della misura e dalla somiglianza tra le firme steady state. Inoltre il riconoscimento degli eventi può essere soggetto a mancati riconoscimenti o all'individuazione di falsi positivi.

Gli approcci state based al contrario non necessitano di sistemi di individuazione degli eventi, utilizzano una state machine, e sono basati sul fatto che quando un elettrodomestico viene acceso o spento o cambia il proprio stato durante il funzionamento crea un insieme di transienti "edge" che hanno una distribuzione di probabilità che corrisponde a quella dell'elettrodomestico esaminato. Questi algoritmi sono basati di solito su Hidden Markov Model

Gli approcci state based sono limitati dalla necessità di fissare a priori lo stato di un elettrodomestico, da lunghi periodi di training, dalla complessità computazionale e dalla difficoltà di trattare stati di accensione molto lunghi.

Si possono distinguere due categorie di features, ossia le misurazioni utili a realizzare la disaggregazione dei carichi:

- transient features: sono caratteristiche del transitorio e sono rappresentate da forma, durata e contenuto armonico del transitorio. Richiedono hardware dedicato ed elevate frequenze per la misura;
- steady state features: includono, potenza attiva e reattiva, la corrente e le forme d'onda di corrente e tensione. Non richiedono hardware dedicato e possono essere ottenute dalle misure standard di uno smart meter. Tra le feature quasi tutti i lavori in letteratura si basano sulla potenza, mentre qualcuno indica la corrente come una misura maggiormente significativa.

Nel processo del NILM si distingue una fase di addestramento dell'algoritmo, e in una fase di classificazione sulla base della lettura. La fase di addestramento o training è la più difficile e maggiormente costosa sia per le risorse impiegate, per le misurazione coinvolte, e per il peso computazionale. Le tecniche di addestramento si dividono in:

- supervised learning;
- unsupervised learning.

La prima tecnica è di gran lunga la più usata, e richiede la misurazione contestuale della potenza consumata dai singoli

elettrodomestici e la potenza assorbita dall'intera utenza domestica con i carichi aggregati. La misurazione dei consumi dei singoli elettrodomestici è funzionale alla classificazione dei carichi. Le tecniche utilizzate sono Support Vector Machines, Nearest Neighbor e diverse forme di Hidden Markov Model. In particolare risulta significativo che possano essere ottenuti risultati accettabili anche con un algoritmo molto semplice come il nearest neighbor, si veda [1].

Le tecniche proposte in letteratura sono moltissime, si possono citare il Discriminative Sparse Coding per permettere la disaggregazione dei carichi in 5 sotto categorie, Factorial Hidden Markov Models in quanto approccio classico per il sequence learning, approcci di tipo Deep Learning tramite Recurrent Neural Networks (RNN), Deep Neural Networks (DNN), Long Short Term Memory (LSTM) e Convolutional Neural Networks (CNN), e Denoising Autoencoders [32] [33] [30] [22].

Un approccio interessante consiste nel subtractive clustering e nella maximum likelihood classification. Le features selezionate per l'identificazione degli elettrodomestici sono il livello di potenza impegnato e la durata del periodo di accensione. I livelli di potenza sono caratterizzati da una distribuzione normale mentre i dati relativi alla durata sono caratterizzati da una distribuzione di Weibull. Si riesce a raggiungere una accuratezza di 86% per l'identificazione di 6 elettrodomestici [27].

Stanno emergendo anche tecniche di tipo unsupervised, che non richiedono una fase di addestramento e non richiedono la misura di utilizzo dei singoli elettrodomestici. Non è richiesta alcuna interazione da parte dell'utente.

Si possono distinguere tre tipologie di metodi unsupervised [48]:

- i metodi che non richiedono alcuna misurazione dei consumi degli elettrodomestici, richiedono però l'inserimento manuale di alcuni dati e non riescono ad essere generalizzati ad utenze diverse
- i metodi che utilizzano un dataset di misure per addestrare un algoritmo che poi viene applicato ad utenze non misurate. La maggior parte dei metodi basati sul deep learning ricade in questa tipologia [31];
- Infine ci sono i metodi che non richiedono alcun addestramento precedente alla fase di addestramento [28].

A. Metodologie di modellazione e trattamento

Le tecniche di raccolta e modellazione dei dati proposte in letteratura sono diverse e variano da paese a paese, in funzione anche della normativa elettrica locale. Nella grande maggioranza dei casi il sistema utilizzato consiste nell'utilizzo di un misuratore di corrente non intrusivo a valle del contatore fiscale, con in aggiunta dei misuratori di corrente per le utenze specifiche utili per l'addestramento dei sistemi di intelligenza artificiale utilizzati per la disaggregazione. Il segnale raccolto dai sensori viene inviato ad un sistema di raccolta e spedito ad un server remoto dove viene conservato ed elaborato. Le caratteristiche specifiche del sistema variano per ogni ricerca esaminata.

E' bene ricordare che l'individuazione dei singoli carichi pone dei seri problemi di privacy dato che le attività degli

utenti possono essere in tal modo facilmente individuate, interpretate e monitorate [11].

In [22] ad esempio, sono proposti due sistemi, una raspberry Pi integrata con un e-Gizmo Power Analyzer capace di misurare carichi fino a 230VAC. Lo stesso Power Analyzer è anche proposto con un Node MCU (<https://it.wikipedia.org/wiki/NodeMCU>) come una soluzione embedded. In entrambi i casi i dati raccolti sono inviati ad un server locale attraverso la LAN ed un router. La Raspberry è utilizzata per valutare il consumo complessivo mentre le NodeMCU sono utilizzati per le smart plug per il consumo dei singoli elettrodomestici. Il server locale interroga raspberry e NodeMCU con cadenza 1s, aggiunge un timestamp al dato e conserva gli stessi in un database PostgreSQL.

La creazione di un training set comporta la pulizia e la creazione di una individual training data per ciascuna utenza, attraverso una ricerca nel database dei punti in cui la utenza in esame è stata attiva. Questi servono per marcare gli istanti in cui una data utenza è stata attiva e per attribuire una etichetta al dato del consumo aggregato.

Accanto alla necessità di raccogliere i dati relativi ai consumi elettrici è ugualmente importante acquisire informazioni circa le necessità dell'utente, anche ai fini dell'addestramento del sistema. Una utenza può essere residenziale o commerciale, in una utenza domestica la composizione della famiglia può rivestire una certa importanza così come la tipologia degli impianti di riscaldamento e raffrescamento e il tipo e la qualità degli elettrodomestici a disposizione. Parimenti la natura dell'attività commerciale riveste un ruolo fondamentale per l'individuazione dei carichi e dei pattern di consumo. L'individuazione di una routine giornaliera e settimanale è in ogni caso un primo passo per l'individuazione di soluzioni personalizzate finalizzate all'ottimizzazione dei consumi energetici.

Una metodologia interessante per il trattamento dei dati è presentata in [4]. Il database analizzato è costituito da tre abitazioni in cui sono misurati il carico aggregato e 6 smart plugs per l'addestramento, questo fa parte di un database di oltre tremila abitazioni. La tecnica applicata all'intero database ha prodotto dei riscontri comparabili al caso di prova nell'80% dei casi, mostrando quindi una buona replicabilità. La procedura può essere divisa in fasi:

- 1) trattamento del segnale per l'eliminazione degli errori di misura ed il filtraggio degli outliers, e per l'eliminazione del rumore della misura;
- 2) individuazione degli eventi di accensione e spegnimento, causati dal variare della potenza impegnata;
- 3) clusterizzazione dei livelli di potenza con un algoritmo k-means;
- 4) generazione delle features, in particolare si selezionano:
 - media e deviazione standard per i valori di potenza dei cluster (vedi punto precedente)
 - media e deviazione standard per la durata degli elementi nei cluster
 - per durate e potenze medie simili si individuano: la media e la deviazione delle fasi ON di un elemento; il numero di occorrenze di eventi in una data finestra

temporale; la distanza temporale media tra eventi successivi dello stesso tipo.

- 5) training del modello attraverso un algoritmo di classificazione degli elettrodomestici basato su Support Vector Machines

B. Benchmarking

Due strumenti molto utili per valutare le performance di un sistema NILM sono il NILMTK (<http://nilmtk.github.io>) che permette di valutare gli algoritmi e ottenere tutte le misure necessarie per la sua caratterizzazione e il prodotto molto simile NILMeval (<https://github.com/beckel/nilm-eval>), il primo è scritto in python, il secondo in Matlab.

C. Librerie

Il già citato NILMTK (<http://nilmtk.github.io>) è un ottimo punto di partenza che permette non solo il trattamento e l'analisi dei dati ma comprende una libreria con alcuni algoritmi di disaggregazione già implementati.

III. TECNICHE DI PROFILAZIONE E CLUSTERING DEGLI UTILIZZATORI ENERGETICI

L'applicazione delle tecniche di data-science ai dati di consumo delle utenze trova diversi ambiti di utilizzo:

- previsione dei consumi
- ottimizzazione della manutenzione degli impianti
- retrofitting di impianti esistenti
- verifica dello stato degli impianti ed individuazioni di eventuali problemi
- analisi economica dei consumi
- individuazione e prevenzione delle frodi

Le tecniche utilizzate per questi ambiti di intervento sono:

- regressione;
- classificazione;
- clustering;
- Association Rules (AR)
- sequence discovery;
- anomaly detection;
- time series analysis

Dei temi precedentemente elencati quello di maggiore interesse per il progetto è certamente quello relativo all'analisi economica dei consumi elettrici. Lo scopo è quello di cercare di capire dall'analisi dei dati, come e quando gli utilizzatori della rete di distribuzione elettrica usano l'energia. Le tecniche utilizzate in quest'ambito sono classificazione, clustering, e pattern analysis.

La conoscenza dei profili di consumo degli utenti rappresenta una risorsa molto importante per i fornitori di energia elettrica, in particolare per quanto riguarda il comportamento nelle peak hours [20].

Uno dei primi studi in quest'ambito è rappresentato dal lavoro di Chicco et al. che hanno raggruppato i clienti in classi sulla base del loro comportamento, questa classificazione rappresentava un primo passo verso la diversificazione delle tariffe [16].

L'idea di base consiste nell'identificare i profili orari di consumo (hourly load profiles, HLP) e nello sviluppare un algoritmo di classificazione automatica dei nuovi utenti [14] [24] [44].

Nei modelli attuali vengono impiegati alcuni parametri per caratterizzare l'utente, come ad esempio il tipo di attività economica, il volume economico dell'attività e il consumo elettrico medio [17].

Le tecniche di classificazione sono generalmente di tipo statistico, come ad esempio k-means [35], tecniche fuzzy [2] e tecniche di riconoscimento dei pattern nel dominio delle frequenze [13].

A. Pre-clustering

La classificazione dei carichi richiede una fase di pre-clustering in cui è necessario individuare diverse macro categorie di utenti. La prima categorizzazione si ha sulla base del contratto sottoscritto e la dimensione economica dell'utenza che è in stretta relazione con il profilo dell'utente.

Altri parametri di interesse per la classificazione, oltre alla dimensione economica è rappresentato dalla variabilità stagionale del consumo elettrico e la tipologia di utenza (domestica, agricola, industriale, servizi pubblici e privati). Ad esempio una utenza agricola è caratterizzata da una forte variabilità stagionale e da una dipendenza dalle condizioni meteorologiche.

Il consumo in genere differisce tra giorni lavorativi e non, con l'eccezione di utenze industriali e commerciali in cui può non esserci distinzione a causa delle caratteristiche del processo produttivo e degli orari di apertura al pubblico.

B. Data reduction

Dopo la fase di pre-clustering segue una fase di data reduction. In genere gli smart meter collezionano le misurazioni di consumo ogni 15min per un totale di 96 letture giornaliere. Queste rappresentano una notevole mole di dati quando vengano moltiplicate per i milioni di utenti di una grande utility. Per questo il dato di consumo giornaliero viene trattato per ridurre la dimensionalità, pur preservandone o anche incrementandone il contenuto informativo, evidenziando cioè le caratteristiche principali del profilo di consumo atte a caratterizzare una tipologia di utente.

Per far questo si utilizzano tecniche come la Principal Component Analysis (PCA) [15], l'analisi armonica [13], e la rappresentazione wavelet del segnale [39].

Per definire dei profili di consumo uniformi tra diverse utenze si cerca di limitare l'influenza della stagionalità raggruppando i dati di consumo in periodi di due o tre mesi. Una prima classificazione può essere fatta già sulla base del consumo medio, ma questo da poche informazioni quando si utilizzi un sistema di tariffazione non flat. Risulta più efficace suddividere separare i dati relativi a giorni lavorativi e non, e separare il consumo per fasce orarie, ad esempio, giornaliera, serale e notturna. Questo però non è sufficiente a descrivere in modo efficace una utenza, in quanto, utilizzare unicamente queste feature porterebbe a raggruppare in un unico cluster utenze che hanno un profilo di consumo flat, ed utenze che

hanno invece un consumo energetico con grandi oscillazioni. Pertanto si aggiunge al consumo medio una misura di questa variabilità espresso ad esempio dal rapporto tra la potenza massima e minima assorbita nel periodo del giorno, e dal numero di picchi e valli nel profilo di consumo nel periodo in esame. Aggiungere ulteriori dettagli porterebbe il numero di informazioni caratterizzanti a crescere tanto da non essere più di immediata comprensione per chi analizza il dato. Selezionare le feature in questo modo è una alternativa alle tecniche automatiche di selezione, come la già citata Principal Component Analysis, e permette di ottenere delle features con un significato fisico di più immediata comprensione.

C. Clustering

La fase di clustering, definisce gruppi di utenti che non sono noti a priori e che richiedono comunque una qualche supervisione per la loro validazione definitiva che viene comunque condotta da parte di esperti del settore.

In letteratura sono disponibili moltissimi algoritmi di clusterizzazione che possono essere divisi in:

- connectivity based
- centroid-based
- distribution based
- density based

Una volta effettuata la clusterizzazione è essenziale validare la suddivisione ottenuta. Questa operazione può essere realizzata attraverso la misura della qualità della clusterizzazione. Le misure riflettono la compattezza, la interconnessione e la separazione tra i cluster, evidenziando quindi se il raggruppamento definisca degli insiemi compatti e ben distinti gli uni dagli altri. Il ricorso a questi indicatori standard permette non solo di selezionare l'algoritmo di clusterizzazione migliore per il task ma anche il numero di cluster ottimale in cui l'insieme di dati può essere scomposto.

Questo processo deve comunque essere supervisionato e guidato da esperti nel settore e dai fruitori finali della clusterizzazione in modo che il numero di raggruppamenti e la loro compattezza possa dar luogo ad una scomposizione che possa essere effettivamente utile.

D. Classificazione

La parte finale del processo è rappresentata dalla classificazione, che permette di assegnare i nuovi utenti a una delle categorie precedentemente definite nella fase di clustering. Anche questa fase viene realizzata in modo automatico attraverso l'algoritmo di classificazione addestrato nella fase precedente, permettendo di trattare una mole di dati altrimenti non gestibile da un singolo operatore. Ogni nuovo utente sarà pertanto assegnato ad una determinata classe. Questo permetterà non solo di prevedere il suo comportamento sulla base dei consumi di utenti "simili" ma anche di individuare eventuali anomalie nei consumi.

In figura 1 [12] è mostrato flusso logico delle operazioni di clustering e classificazione.

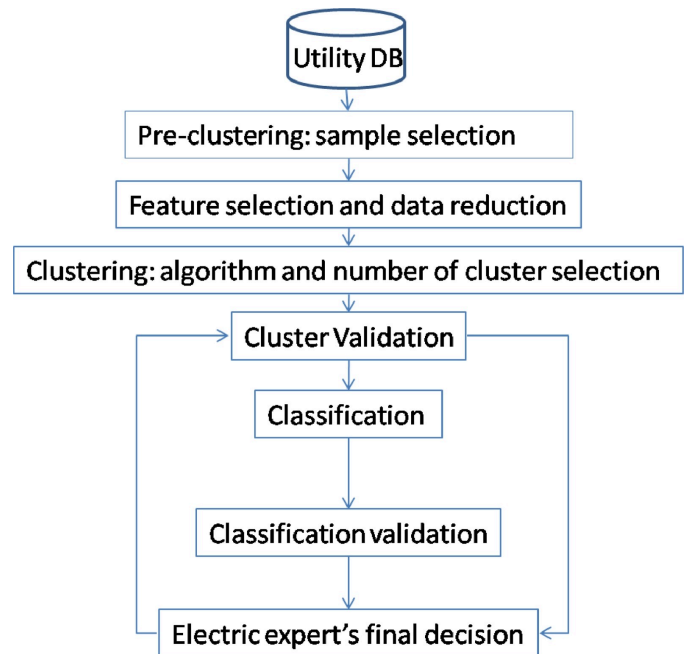


Figura 1. Framework per clustering e classificazione

IV. DATASET PUBBLICI

La disponibilità di dataset con misurazioni di dati aggregati e dei consumi dei singoli elettrodomestici è essenziale, permette di lavorare su dati già certificati e permette di confrontare le performance degli algoritmi sviluppati con quelli di altri ricercatori nello stesso tema.

Nella tabella seguente tratta da [19] sono riportati i più importanti dataset oggi disponibili in letteratura e liberamente utilizzabili.

Le referenze sono [34] [37] [21] [3] [47] [45] [38] [31] [7] [41] [40] [8] [23].

RIFERIMENTI BIBLIOGRAFICI

- [1] Hana Altrabalsi, Jing Liao, Lina Stankovic, and Vladimir Stankovic. A low-complexity energy disaggregation method: Performance and robustness. In *Computational Intelligence Applications in Smart Grid (CIASG), 2014 IEEE Symposium on*, pages 1–8. IEEE, 2014.
- [2] A Azadeh, M Saberi, and O Seraj. An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: a case study of iran. *Energy*, 35(6):2351–2366, 2010.
- [3] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. Smart*: An open data set and tools for enabling research in sustainable homes. *SustKDD, August*, 111:112, 2012.
- [4] Kaustav Basu, Vincent Debusschere, Seddik Bacha, Ahmad Hably, Danny Van Delft, and Geert Jan Dirven. A generic data driven approach for low sampling load disaggregation. *Sustainable Energy, Grids and Networks*, 9:118–127, 2017.
- [5] Kaustav Basu, Vincent Debusschere, Seddik Bacha, Ujjwal Maulik, and Sanghamitra Bondyopadhyay. Nonintrusive load monitoring: A temporal multilabel classification approach. *IEEE Transactions on Industrial Informatics*, 11(1):262–270, 2015.
- [6] Kaustav Basu, Vincent Debusschere, Ahlame Douzal-Chouakria, and Seddik Bacha. Time series distance-based methods for non-intrusive load monitoring in residential buildings. *Energy and Buildings*, 96:109–117, 2015.
- [7] Nipun Batra, Haimonti Dutta, and Amarjeet Singh. Indic: Improved non-intrusive load monitoring using load division and calibration. In *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, volume 1, pages 79–84. IEEE, 2013.

Dataset	Location	Duration	No. of houses	Sensors/house	Resolution	Features	Other Data
REDD [62]	USA	3-19 days	6	24	15KHz(Aggr), 0.5Hz and 1Hz (sub)	V and P (Aggr), P (sub)	
BERDS [72]	USA	1 year	1	4	20sec	P,Q and S	climate data
BLUED [73]	USA	8 days	1	Aggregated	12KHz(Aggr only)	I, V and State transition label for each appliance.	
Smart [74]	USA	3 months	3	21-26 circuit meters	1Hz	P and S (Aggr), P (Sub)	electricity generation data from on-site solar panels and wind turbines, outdoor weather data, temperature and humidity data in indoor rooms
DRED [75]	Netherlands	6 months	3	12 appliances	1Hz	P (Aggr & Sub)	indoor temperature, outside temperature, wind speed, pre-cipitation, humidity and occupancy data.
Tracebase [76]	Germany	N/A	15	158 devices	1-10sec(Sub only)	P	
AMPDS [28]	Canada	1 year	1	19	1min	V, I, F P, Q, S and Pf	water and natural gas,
AMPds2 [77]	Canada	2	1	21	1min	V, I, F P, Q, S, Pf, real energy, reactive energy, and apparent energy	water and natural gas, weather data and utility billing data.
UK-DALE [69]	UK	499 days, years(house 1)	2.5	5	5-54 devices	16 kHz(Aggr) and 1/6 Hz(Sub)	P and switch status
iAWE [59]	India	73 days	10	33 devices	1sec(Aggr) and 1sec or 6sec (Sub)	V, I, F, P and phase	Water and ambient conditions
REFIT [78]	UK	2years	20	11	8sec	P	Gas and environmental data
GREEND [79]	Austria/ Italy	1year	9	9	1Hz	P	
ECO [4]	Switzerland	8months	6		1Hz	P and Q	Occupancy information
IHEPCDS ⁹	France	4 years	1	3	1min	V, I, P and Q	
OCTES ¹⁰	Scotland,Iceland &Finland	413months	33	Aggregated	7sec	P and phase	
HES	UK	1month(255 houses), 1year(26houses)	251	13-51	2min	P	
ACS-F1 [80]	Switzerland	2, 1 hoursessions	NA	100, 10 types	10sec	P, Q, I, f, V and phase	

Aggregate (Aggr), Sub-metering (sub), Active Power (P), Reactive Power (Q), Apparent Power (S), Energy (E), Frequency (f), Voltage (V) and Current (I)

Figura 2. Principali dataset pubblici per la disaggregazione dei carichi elettrici

- [8] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, pages 80–89. ACM, 2014.
- [9] Mario E Berges, Ethan Goldman, H Scott Matthews, and Lucio Soibelman. Enhancing electricity audits in residential buildings with non-intrusive load monitoring. *Journal of industrial ecology*, 14(5):844–858, 2010.
- [10] Thomas Bier, Dirk Benyoucef, Djaffar Ould Abdeslam, Jean Merckle, and Philipp Klein. Smart meter systems measurements for the verification of the detection & classification algorithms. In *Industrial Electronics Society, IECON 2013-39th Annual Conference of the IEEE*, pages 5000–5005. IEEE, 2013.
- [11] Benjamin J Birt, Guy R Newsham, Ian Beausoleil-Morrison, Marianne M Armstrong, Neil Saldanha, and Ian H Rowlands. Disaggregating categories of electrical energy end-use from whole-house hourly data. *Energy and Buildings*, 50:93–102, 2012.
- [12] Félix Biscarri, Iñigo Monedero, Antonio García, Juan Ignacio Guerrero, and Carlos León. Electricity clustering framework for automatic classification of customer loads. *Expert Systems with Applications*, 2017.
- [13] Enrico Carpaneto, Gianfranco Chicco, Roberto Napoli, and Mircea Scutariu. Electricity customer classification using frequency-domain load pattern data. *International Journal of Electrical Power & Energy Systems*, 28(1):13–20, 2006.
- [14] Gianfranco Chicco. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42(1):68–80, 2012.
- [15] Gianfranco Chicco, Roberto Napoli, and Federico Piglion. Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, 21(2):933–940, 2006.
- [16] Gianfranco Chicco, Roberto Napoli, Federico Piglion, Petru Postolache, Mircea Scutariu, and Cornel Toader. Load pattern-based classification of electricity customers. *IEEE Transactions on Power Systems*, 19(2):1232–1239, 2004.
- [17] O Dzobo, Karin Alvehag, CT Gaunt, and R Herman. Multi-dimensional customer segmentation model for power system reliability-worth analysis. *International Journal of Electrical Power & Energy Systems*, 62:532–539, 2014.
- [18] Ehsan Elhamifar and Shankar Sastry. Energy disaggregation via learning powerlets and sparse coding. In *AAAI*, pages 629–635, 2015.
- [19] Anthony Faustine, Nerey Henry Mvungi, Shubi Kaijage, and Kisan-giri Michael. A survey on non-intrusive load monitoring methodologies and techniques for energy disaggregation problem. *arXiv preprint arXiv:1703.00785*, 2017.
- [20] Adonias Magdiel Silva Ferreira, Cristiano Hora de Oliveira Fontes, Carlos Arthur Mattos Teixeira Cavalcante, and Jorge Eduardo Soto Marambio. Pattern recognition as a tool to support decision making in the management of the electric sector. part ii: A new method based on clustering of multivariate time series. *International Journal of Electrical Power & Energy Systems*, 67:613–626, 2015.
- [21] Adrian Filip. Blued: A fully labeled public dataset for event-based non-intrusive load monitoring research. In *2nd Workshop on Data Mining Applications in Sustainability (SustKDD)*, page 2012, 2011.
- [22] Felan Carlo C Garcia, Christine May C Creayla, and Erees Queen B Macabebe. Development of an intelligent system for smart home energy disaggregation using stacked denoising autoencoders. *Procedia Computer Science*, 105:248–255, 2017.
- [23] Christophe Gislser, Antonio Ridi, Damien Zufferey, Omar Abou Khaled, and Jean Hennebert. Appliance consumption signature database and recognition test protocols. In *Systems, Signal Processing and their Applications (WoSSPA), 2013 8th International Workshop on*, pages 336–341. IEEE, 2013.
- [24] Gheorghe Grigoras and Florina Scarlatache. Knowledge extraction from smart meters for consumer classification. In *Electrical and Power Engineering (EPE), 2014 International Conference and Exposition on*, pages 978–982. IEEE, 2014.
- [25] Sidhant Gupta, Matthew S Reynolds, and Shwetak N Patel. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139–148. ACM, 2010.
- [26] George William Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.
- [27] Nilson Henao, Kodjo Agbossou, Souso Kelouwani, Yves Dubé, and Michaël Fournier. Approach in nonintrusive type i load monitoring using subtractive clustering. *IEEE Transactions on Smart Grid*, 8(2):812–821, 2017.
- [28] Ruoxi Jia, Yang Gao, and Costas J Spanos. A fully unsupervised non-intrusive load monitoring framework. In *Smart Grid Communications*

- (*SmartGridComm*), 2015 *IEEE International Conference on*, pages 872–878. IEEE, 2015.
- [29] Georgios Kalogridis, Costas Efthymiou, Stojan Z Denic, Tim A Lewis, and Rafael Cepeda. Privacy for smart meters: Towards undetectable appliance load signatures. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pages 232–237. IEEE, 2010.
- [30] Jack Kelly and William Knottenbelt. Neural nilm: Deep neural networks applied to energy disaggregation. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 55–64. ACM, 2015.
- [31] Jack Kelly and William Knottenbelt. The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific data*, 2:150007, 2015.
- [32] J Zico Kolter, Siddharth Batra, and Andrew Y Ng. Energy disaggregation via discriminative sparse coding. In *Advances in Neural Information Processing Systems*, pages 1153–1161, 2010.
- [33] J Zico Kolter and Tommi Jaakkola. Approximate inference in additive factorial hmms with application to energy disaggregation. In *Artificial Intelligence and Statistics*, pages 1472–1482, 2012.
- [34] J Zico Kolter and Matthew J Johnson. Redd: A public data set for energy disaggregation research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, volume 25, pages 59–62, 2011.
- [35] José J López, José A Aguado, F Martín, F Muñoz, A Rodríguez, and José E Ruiz. Hopfield-k-means clustering algorithm: A proposal for the segmentation of electricity customers. *Electric Power Systems Research*, 81(2):716–724, 2011.
- [36] Jean-Nicolas Louis, Antonio Caló, Kauko Leiviskä, and Eva Pongrácz. Modelling home electricity management for sustainability: The impact of response levels, technological deployment & occupancy. *Energy and Buildings*, 119:218–232, 2016.
- [37] Mehdi Maasoumy, B Sanandaji, Kameshwar Poolla, and Alberto Sangiovanni Vincentelli. Berds-berkeley energy disaggregation data set. In *Proceedings of the Workshop on Big Learning at the Conference on Neural Information Processing Systems (NIPS)*, 2013.
- [38] Stephen Makonin, Bradley Ellert, Ivan V Bajić, and Fred Popowich. Electricity, water, and natural gas consumption of a residential house in canada from 2012 to 2014. *Scientific data*, 3, 2016.
- [39] Stephane G Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7):674–693, 1989.
- [40] Andrea Monacchi, Dominik Egarter, Wilfried Elmenreich, Salvatore D’Alessandro, and Andrea M Tonello. Greend: an energy consumption dataset of households in italy and austria. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*, pages 511–516. IEEE, 2014.
- [41] David Murray, Lina Stankovic, and Vladimir Stankovic. An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific data*, 4:160122, 2017.
- [42] Peter Palensky and Dietmar Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE transactions on industrial informatics*, 7(3):381–388, 2011.
- [43] A Prudenzi. A neuron nets based procedure for identifying domestic appliances pattern-of-use from energy recordings at meter panel. In *Power Engineering Society Winter Meeting, 2002. IEEE*, volume 2, pages 941–946. IEEE, 2002.
- [44] Sérgio Ramos, João M Duarte, F Jorge Duarte, and Zita Vale. A data-mining-based methodology to support mv electricity customers’ characterization. *Energy and Buildings*, 91:16–25, 2015.
- [45] Andreas Reinhardt, Paul Baumann, Daniel Burgstahler, Matthias Hollick, Hristo Chonov, Marc Werner, and Ralf Steinmetz. On the accuracy of appliance identification based on distributed load metering data. In *Sustainable Internet and ICT for Sustainability (SustainIT), 2012*, pages 1–9. IEEE, 2012.
- [46] Thomas Strasser, Filip Andrén, Munir Merdan, and Alexander Prostejovsky. Review of trends and challenges in smart grids: An automation point of view. In *International Conference on Industrial Applications of Holonic and Multi-Agent Systems*, pages 1–12. Springer, 2013.
- [47] Akshay SN Uttama Nambi, Antonio Reyes Lua, and Venkatesha R Prasad. Locoed: Location-aware energy disaggregation framework. In *Proceedings of the 2Nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 45–54. ACM, 2015.
- [48] Bochao Zhao, Lina Stankovic, and Vladimir Stankovic. On a training-less solution for non-intrusive appliance load monitoring using graph signal processing. *IEEE Access*, 4:1784–1799, 2016.