International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:12, No:1, 2018

Water End-Use Classification with Contemporaneous Water-Energy Data and Deep Learning Network

Khoi A. Nguyen, Rodney A. Stewart, Hong Zhang

***Abstract*—'Water-related energy' is energy use which is directly or indirectly influenced by changes to water use. Informatics applying a range of mathematical, statistical and rule-based approaches can be used to reveal important information on demand from the available data provided at second, minute or hourly intervals. This study aims to combine these two concepts to improve the current water end use disaggregation problem through applying a wide range of most advanced pattern recognition techniques to analyse the concurrent high-resolution water-energy consumption data. The obtained results have shown that recognition accuracies of all end-uses have significantly increased, especially for mechanised categories, including clothes washer, dishwasher and evaporative air cooler where over 95% of events were correctly classified.

***Keywords*—**Deep learning network, smart metering, water end use, water-energy data.

I. Introduction

EMERGING technologies and the associated big data informatics, once fully understood and exploited, are the truly "smart" components of a digital water and electricity grid, and these informatics can be used for a wealth of applications [1]-[5]. Intelligent metering uptake however, remains relatively slow, due largely to the unexploited benefits from the back-end of the smart grid, including meters and sensors. The water-energy nexus can be considered as the "interconnections" or "cause-and-effect" relationships between the two resources. In the context of this study, contemporaneous water-energy data is important to understand the water-energy nexus within households, industry and commerce. As energy use related to water use within households, industry and commerce typically accounts for more than 80% of the energy use in the "urban water cycle" [6]-[9], the main aims of this paper are to firstly introduce a new embedded system for collecting high resolution residential water and energy consumption in almost real-time manner, and then use this contemporaneous data to improve the task of water end-use disaggregation. To address this issue, a range of mathematical, statistical and rule-based approaches were all combined, including Artificial Deep Learning Network, Dynamic Time Warping Algorithm, Hidden Markov model, Signal Filtering and Time-of-day Probability. The next section will present in detail how a new embedded data collection kit was set up to record concurrent water – energy data every five seconds, which is followed by the development of a new method to analyse the collected concurrent data to enhance water end-use disaggregation accuracy.

Khoi. A. Nguyen is with Griffith University, Australia (corresponding author, e-mail: k.nguyen@griffith.edu.au).
Rodney A. Stewart and Hong Zhang are with Griffith University, Australia (e-mail: r.stewart@griffith.edu.au, hong.zhang@griffith.edu.au).

II. High Resolution Water-Energy Data Collection

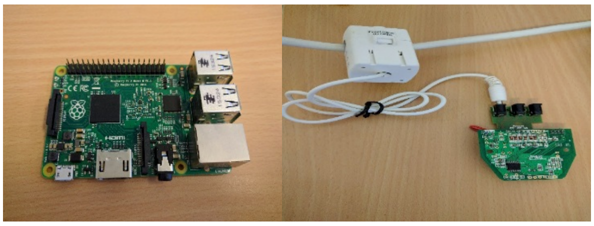
This section describes how a high-resolution data collection kit for both water and energy data was developed and trialed in this study. For energy data collection, three different devices including a current transformer, a data transmitter and a microprocessor board (i.e. Raspberry pi) were employed as follows:

Step 1. A current transformer (CT) was firstly clipped over a live electricity cable to produce an alternating current in its secondary winding which is proportional to the current being measured in its primary. Current transformer reduces high voltage currents to a much lower value and provide a convenient way of safely monitoring the actual electrical current flowing in an AC transmission line using a standard meter.

Step 2. An Arduino board was used to convert the current transformer output into radio signal and send it to a receiver every 5 seconds (Fig. 1 (a)).

Step 3. A receiver (i.e. Raspberry pi processor in this case) was used to collect the signal transmitted from Step 2. At this stage, an application written in C and Python programming language was used to decode the data to facilitate for later analysis (Fig. 1 (b)). In terms of water consumption data, a commercial smart water meter (Fig. 1 (c)) was used to record consumption data every 5 seconds. An application written in Python language was developed and embedded into the Raspberry pi (Fig. 1 (b)) to allow it to wirelessly connect to smart water meter.

Step 4. The Raspberry pi processor in previous step is then connected to an external display to show both concurrent water – energy consumption in a real time manner (Fig. 1 (d)).



(a) CT and transmitter (b) Raspberry pi processor

(c) Smart water meter (d) Raspberry pi with display

Fig. 1 Real-time water-energy consumption data collection

III. Water End-Use Disaggregation with Concurrent Data

With the collected contemporaneous water – energy data, the disaggregation process of water end-use events was developed following the procedure presented in Fig. 2.

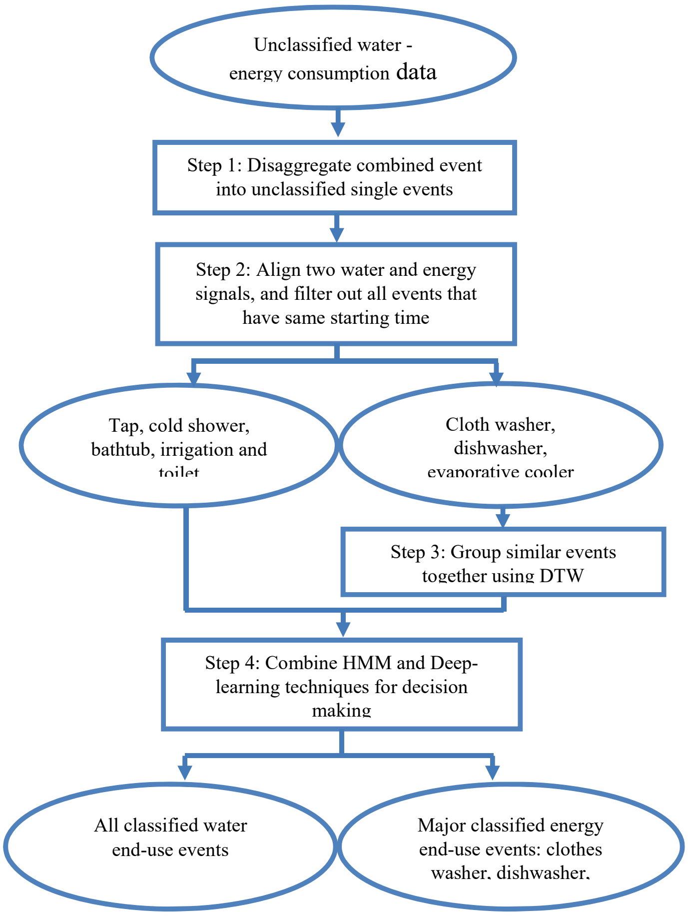


Fig. 2 Water end-use disaggregation process

Step 1. Disaggregation of a combined event into isolated unclassified single events

Due to the fact that in any household, there is always more than one electrical device operating at the same, the obtained energy data is the aggregation of power consumption from several concurrent devices. To make use of this energy signal, the very first step is to disaggregate the overall combined event into all single unclassified events (Fig. 3) using the following logic:

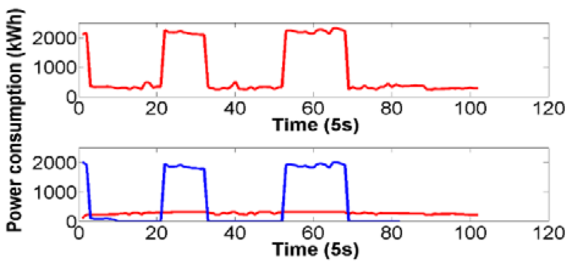


Fig. 3 Separation of a combined event into single events

The energy data records the power usage; therefore, the power changes (gradients) indicate a device is switched on or off. A modified gradient vector filtering method is developed to allow the analyst to dissect a combined event to any desired level. The principle of this technique is based on an examination of the gradient change, alongside the event sample, to make different levelling decisions. The method derivation can be explained as follows:

Given a combined event sample, whose power is demonstrated as a vector $\boldsymbol{a} = (a_1, a_2, \ldots a_j \ldots , a_m)$ of length m, the gradient of vector $\boldsymbol{a}$, $\boldsymbol{g} = (g_1, g_2, \ldots g_j \ldots , g_{m-1})$, is defined as:

$$g_j = \frac{(a_{j+1} - a_j)}{dt}, 1 \leq j < m \quad (1)$$

where j is the sampling index, a_j is power, expressed in terms of kW recorded at time t_j, and dt is the sampling interval ($dt = t_{j+1} - t_j$). The increase of power gradient indicates the occurrence of a new power use, while the decrease of power implies the completion of a power use. Based on that, the starting and ending moment of each single event can be determine to allow of the separation to be undertaken.

Step 2. Identification of mechanised end use events

Once all single water and energy events have been isolated, the next step is two align these two signals to find all events that start at the same time. This step aims to separate all mechanised end use categories including clothes washer, dishwasher, evaporative air cooler and instantaneous hot shower system from the remaining categories (including shower, toilet, tap and irrigation) with the accuracy of almost 100%. Presented in Fig. 4 below is a group of three water and energy events that have the same starting and ending time.

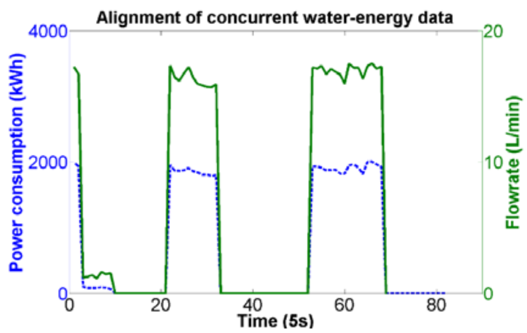


Fig. 4 Alignment of concurrent water-energy data

Step 3. Grouping of all similar events together using Dynamic Time Warp algorithm (DTW)

Once all water and energy events having identical starting time have been successfully separated in Step 2, the next step is to group them together based on their pattern similarity. This process can be achieved using Dynamic Time Warping algorithm as presented in [10]. At the end of this process, there will be several groups whose events belong to one particular category. The following step aims to identify the correct end-use category for all unclassified events and groups.

Step 4. Water end use disaggregation using Deep learning and Hidden Markov model (HMM)

This step aims at combining the Stacked Auto-encoder network [11]-[13] with HMM to assign all unclassified events (containing shower, tap, toilet, irrigation and bathtub), and unclassified groups (containing clothes washer, dishwasher, evaporative air cooler and hot instantaneous hot shower if there is) into appropriate end-use categories. The use of Hidden Markov Model (HMM) in the task of water end-use classification have been presented in [14], [15], but this is the first time HMM is combined with Deep learning technique and energy data to enhance the water end-use disaggregation process. The architecture of the applied deep learning network is presented below which includes the following characteristics:

- Input data for training: 196550 events from all categories
- Encoder 1 (unsupervised training) : 150 hidden layers with maximum 500 training iterations
- Encoder 2 (unsupervised training) : 100 hidden layers with maximum 500 training iterations
- Encoder 3 (unsupervised training) : 50 hidden layers with maximum 500 training iterations
- Final layer (supervised training) : 500 training iterations
- Output layer: Eight different end use categories including: shower, tap, clothes washer, dishwasher, evaporative air cooler, bathtub, toilet, and irrigation.

At the end of this training process using stacked auto-encoder technique, a deep learning network will be obtained. In terms of HMM model development, the same number of input data will also be used to develop eight different HMM models for eight end use categories mentioned above. The HMM training algorithm for water end use classification is presented in the Appendix.

With the availability of the customised HMM and deep-learning model, the likelihood estimation of each event using these two techniques is conducted following the below logic: Given $\boldsymbol{A} = (a_1, a_2, \ldots, a_8)$ is the obtained probability according to which an unknown event will be classified as shower (a_1), faucet (a_2), clothes washer (a_3), dishwasher (a_4), toilet (a_5), bathtub (a_6), irrigation (a_7) and evaporative air cooler (a_8) using HMM, $\boldsymbol{B} = (b_1, b_2, \ldots, b_8)$ being the achieved likelihood when estimating this event with deep-learning model, and $\boldsymbol{C} = (c_1, c_2 \ldots c_i, \ldots c_8)$ being the combined likelihood of A and B, where c_i is calculated using (2) and ($1 \leq i \leq 8$), then an unclassified event will be assigned to category (i) if c_i is the maximum value of $\boldsymbol{C}$.

$$c_i = (1 - e^{1/(a_i)})\, b_i \quad (2)$$

For a group $\boldsymbol{M}$ containing n events of the same category, denoted as $\boldsymbol{M} = (m_1, m_2, \ldots, m_n)$, the aggregate event likelihood of this group to be classified as category, denoted as $\boldsymbol{M}c_i$, will be determined as follows:

$$\boldsymbol{M}c_i = \frac{\sum_{j=1}^{n} c_{ij}}{n} \quad (3)$$

where c_{ij} is the aggregate likelihood of event j in the group to be assigned to category i. By following this process, all events in a group will be assigned to category i if:

$$i = \text{argmax}_{1 \le i \le 8}(\boldsymbol{M}c_i) \tag{4}$$

IV. Comparison with Other Conventional Approaches

The proposed method has helped boost the achieved recognition accuracy of mechanised end-use, including clothes washer, dishwasher and evaporative air cooler to almost absolute, as presented in Table I. The most significant contribution of this approach is at addressing the problem with evaporative air cooler, which is fairly complicated and not popular in many regions in the world. The approach using water consumption data only in combination with HMM and ANN has resulted the maximum accuracy of 73.8% due to the inefficiency of input data for training. In this study, the concurrent water-energy signal has helped significantly reduce the problem difficulty by isolating the main mechanised events, including evaporative coolers from the other end use categories at the first step of the overall process. As a result, the obtained accuracy for this end-use was raised from 73.8% to 95.1%. Clothes washer and dishwasher classification accuracies also show the same trend when they increased from 91.7% to 97.8% and 96% to 98.2%, respectively.

TABLE I
Accuracy Comparison with Other Approaches

	No. of samples for testing	Recognition accuracy (%)			**Water + energy data + HMM + deep learning**
		HMM	ANN	HMM+ANN	
Shower	3,000	70.3	78.6	93.8	93.8
Faucet	3,000	76.2	71.7	90.8	90.8
C. washer	3,000	82.6	72.8	91.7	**97.8**
Dishwasher	3,000	90.9	76.9	96.0	**98.2**
Toilet	3,000	89.4	88.8	94.4	94.4
Bathtub	500	64.0	88.0	88.1	89.1
Irrigation	500	63.6	70.6	85.9	86.1
Air cooler	**500**	**62.5**	**68.1**	**73.8**	**95.1**
Overall	16,000	80.8	77.9	90.4	93.1

V. Conclusion and Future Study

The establishment of an integrated water management system, which employs smart water metering, in conjunction with a series of intelligent algorithms to automate the flow trace analysis process, is becoming more and more popular. However, due to technology constraints and data inefficiency, the obtained accuracies of most available water end use disaggregation systems now have not reached the utilities' expectation to be widely implemented. This study has brought this research area into the next level through utilising water and energy data, in combination with the most advanced deep-learning neural network to significantly boost the obtained accuracy of some major categories to above 95%. Moreover, the proposed approach also helped classify some of main residential energy end-uses including hot water, clothes washer, dishwasher and evaporative air cooler. Future research will focus on using water consumption data to enhance the process of energy end-use disaggregation which still remains problematic due to the lack of input high resolution data for investigation.

Appendix

In the training and classification process with HMM, flow rate series of each event is the only required input as this technique relies on the analysis of flow rate change along the event to make different decisions. The measured water flow sequence is defined as $\mathbf{O} = (o_1, o_2, \dots o_t, \dots, o_T)$, where t is the observation time index and T is the total number of flow observations; a state vector is defined as $\mathbf{q} = (1, 2, 3 \dots i, \dots, N)$, where i is the state at index i and N is the maximum or the last state; the initial state probability is π_j; the state transition probability is a_{ij}; and the observation probability $b_j(o_k)$; where i and j are the state indices. Training algorithms of the HMM models for this study are summarised below using (5)-(12):

Step 1. Assume random probabilities for initial state probability π_j, state transition probability a_{ij}, and observation probability $b_j(o_k)$ as their initial values *with the restrictions as follows:*

$\sum_{j=1}^{N_s} \pi_j = 1$: Total probabilities of starting in state j at time t_1 is equal to 1. It should be noted from [14] that the number of state, N_s, was tested from 1 to 400, and it was found that the HMM model established using N_s = 100 yields the highest recognition accuracy; therefore, a random vector containing 100 values, whose sum is 1, was selected for initial state probability π_i

$\sum_{j=1}^{N_s} a_{ij} = 1$: Total transition probability from state i to all other states is equal to 1; therefore, random matrix with N_s rows and N_s columns, whose sum of each row is 1, was selected for state transition probability a_{ij}

$\sum_{k=1}^{N_p} b_j(o_k) = 1$: Total probability of having observation o_k at state j is equal to 1. N_p is the number of the possible observations as proposed for this study (i.e. the maximum flow rate recorded from water meter of any residential household never exceeds $N_p = 300$ pulses), a matrix with N_s rows and N_p columns, whose sum of each column is 1, is randomly selected for observation probability $b_j(o_k)$

Step 2. Using the values from Step 1, determine the following parameters:

- $\alpha_t(i)$: the probability of flow rate o_1 through to o_t and being in state i at time t, ($q_t = i$), given the HMM (λ)

$$\alpha_t(i) = P(o_1 o_2 \dots o_t, q_t = i \mid \lambda) \tag{5}$$

- $\beta_t(i)$: the probability of flow rate o_{t+1} through to o_T, given the HMM (λ) and given that the model is currently in state i at time t ($q_t = i$)

$$\beta_t(i) = P(o_{t+1}\, o_{t+2} \cdots o_T \mid q_t = i, \lambda) \tag{6}$$

- $\gamma_t(i)$: the probability of being in state i at time t given a water flow sequence ($\mathbf{O}$) and HMM (λ).

$$\gamma_t(i)=\frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)} \tag{7}$$

- $\xi_t(i,j)$: the probability of being in state i at time t, and in state j at time $t+1$, given a water flow sequence ($\mathbf{O}$) and the HMM (λ) as:

$$\xi_t(i,j)=\frac{P(q_t=i,q_{t+1}=j,\mathbf{O}\mid\lambda)}{P(\mathbf{O}\mid\lambda)} \tag{8}$$

where

$$P(\mathbf{O}\mid\lambda)=\sum_{k=1}^{N}\sum_{p=1}^{N}\alpha_t(k)a_{kp}b_p(o_{t+1})\beta_{t+1}(p),$$

$$P(q_t=i,q_{t+1}=j,\mathbf{O}\mid\lambda)=\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)$$

Step 3. Calculate the following parameters for each water flow sequence (**O**)

$\sum_{t=1}^{T}\gamma_t(i)$: expected number of times in state i for the water flow sequence (**O**).

$\sum_{t=1}^{T-1}\gamma_t(i)$: expected number of transition from state i for the water flow sequence (**O**).

$\sum_{t=1}^{T-1}\xi_t(i,j)$: expected number of transition from state i to state j for the flow rate sequence (**O**).

Step 4. With the calculated values in Step 3, the probabilities values of π_i, a_{ij} and $b_j(o_k)$ can be updated by performing (9)-(11):

$$\overline{\pi}_i=\gamma_1(i) \tag{9}$$

$$\overline{a}_{ij}=\frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \tag{10}$$

$$\overline{b}_j(o_k)=\frac{\sum_{\substack{t=1\\ \text{such that } \mathbf{o}_t=\mathbf{v}_k}}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)} \tag{11}$$

It should be noted that the above calculations of π_j, a_{ij} and $b_j(o_k)$ will be updated for any new water flow rate sequence **(O)** introduced into the HMM. The overall process is completed when all samples have been provided for training. Once the final HMM (λ) with π_j, a_{ij} and $b_j(o_k)$ are available, the recognition process can be applied for any new water flow sequence. The probability of a water flow sequence given a HMM can be determined using the following formula:

$$P(\mathbf{O}\mid\lambda)=\sum_{i=1}^{N}\alpha_T(i) \tag{12}$$

where $\alpha_T(i)$ is the probability of flow rate $o_1$through to o_T and being in state i at time T, given the HMM (λ).

ACKNOWLEDGMENT

Project support is acknowledged from the Australian Research Council (ARC) Linkage Project (LP160100215) that also includes funding from the following industry partners: Yarra Valley Water, City West Water, Southeast Water and Aquiba. The University of Technology (UTS), Sydney is also a partnered investigator to this grant. Special thanks to Griffith eResearch Services Team for developing and supporting the Smart Meter Information Portal. Finally, the authors greatly appreciate the kind support of the households involved in the study.

REFERENCES

[1] Klein, G., Krebs, M., Hall, V., O'Brien, T. & Blevins, B. B., (2005). *California's water– energy relationship. Final Staff Report.* California Energy Commission.
[2] Nguyen, K. A., Stewart, R. A., Zhang, H., & Jones, C. (2015). Intelligent autonomous system for residential water end use classification: Autoflow. *Applied Soft Computing, 31*, 118-131.
[3] Makki, A. A., Stewart, R. A., Beal, C. D., & Panuwatwanich, K. (2015). Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption. *Resources, Conservation and Recycling, 95*, 15-37.
[4] Stewart, R., Giurco, D., & Beal, C. (2013). Age of intelligent metering and big data: Hydroinformatics challenges and opportunities. *Journal of the International Association for Hydro-environment Engineering and Research, 2*, 107-110.
[5] Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews, 56*, 215-225.
[6] Kenway, S. J., Binks, A., Lane, J., Lant, P. A., Lam, K. L., & Simms, A. (2015). A systemic framework and analysis of urban water energy. *Environmental Modelling and Software, 73*, 272-285.
[7] Arpke, A., & Hutzler, N. (2006). Domestic Water Use in the United States. A Life-Cycle Approach. *Journal of Industrial Ecology, 10*(1-2), 169-183.
[8] Cheng, C. L. (2002). Study of the inter-relationship between water use and energy conservation for a building. *Energy and Buildings, 34*(3), 261-266.
[9] Stokes, J.R., & Horvath, A. (2009). Energy and Air Emissions Impacts of Water Supply. *Environmental Science and Technology, 43*(8), 2680–2687J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility (Periodical style)," *WASET Trans. Electron Devices*, vol. ED-11, pp. 34–39, Jan. 1959.
[10] Nguyen, K. A., Zhang, H., & Stewart, R. A. (2011). Application of dynamic time warping algorithm in prototype selection for the disaggregation of domestic water flow data into end use events.

Proceeding of the 34th World Congress of the International Association for Hydro-Environment Engineering and Research, Brisbane, Australia.

[11] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A (2010) Stacked denosing Autoencoders: Learning Representations in a Deep Network with Local Denoising Criterion. *Journal of Machine Learning Research*, 11, 3371-3408.

[12] Weston, J., Ratle, F., and Collobert, R. Deep learning via semi-supervised embedding. In William W. Cohen, Andrew McCallum, and Sam T. Roweis, editors, Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML'08), pages 1168–1175, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-205-4. doi: 10.1145/1390156.1390303.

[13] Chapelle, O., Scholkopf, B., and Zien, A. editors. Semi-Supervised Learning. MIT Press, Cambridge, MA, 2006.

[14] Nguyen, K. A., Zhang, H., & Stewart, R. A. (2013a). Intelligent pattern recognition model to automate the categorisation of residential water end-use events. *Journal of Environmental Modelling and Software, 47*, 108-127.

[15] Nguyen, K. A., Zhang, H., & Stewart, R. A. (2013b). Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-environment Research*, *7*(3), 182-201.