City University of New York (CUNY)
CUNY Academic Works

International Conference on Hydroinformatics

2014

Development Of An Autonomous And Intelligent System For Residential Water End Use Classification

Khoi Anh Nguyen

Rodney Stewart

Hong Zhang

How does access to this work benefit you? Let us know!

More information about this work at: https://academicworks.cuny.edu/cc_conf_hic/8
Discover additional works at: https://academicworks.cuny.edu

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

HYBRID ANN AND HMM BASED INTELLIGENT RESIDENTIAL WATER END-USE CLASSIFICATION SYSTEM

KHOI A. NGUYEN, RODNEY A. STEWART, HONG ZHANG, ANTHONY BROWN
[1]*Griffith University, School of Engineering, Parkland Drive, QLD 4215, Australia*
[2]*Yarra Valley Water, Lucknow Street, Mitcham, VIC 3132, Australia*

Abstract
Greater than 50% of the world's population now live in urban areas and this proportion is projected to grow further in the coming decades (Grimm et al. 2008). The advanced management of urban water consumption is essential to a sustainable water future. Understanding how, when and why water is used is useful for helping to manage water supply and demand in urban water system. Detailed information on the quantities of water used in different circumstances is critical to forecasting the demand for water and planning for water security. Current intelligent water metering systems allow for the high resolution time series recording of water consumption. Such a high resolution of data is necessary to classify flow patterns into each and every water end-use event in the household in a particular recording period (i.e. tap, clothes washer, shower, etc.). However, the present water end-use classification techniques are both resource intensive and limited in accuracy. To overcome these deficiencies, this research developed an autonomous and intelligent system for residential water end-use classification using the Hidden Markov Model (HMM) coupled with Artificial Neural Networks (ANN). Large water end use datasets from Brisbane and Melbourne in Australia were used to train and validate the system. The hybrid model's water end use category recognition accuracy ranged from 85.9 to 96% which is a significant improvement over existing methods.

1. Introduction

Intelligent metering technology coupled with advanced numerical techniques enables a paradigm shift in the water information available to the customer and water business. This paper introduces a new approach to the problem of autonomous domestic water end use classification and a concept for building an integrated system based on a dynamic and user-friendly communication environment.

The key enabler for this system is the development of pattern matching algorithms which are able to automatically categorise high resolution flow data into particular water end-use categories. There are currently three approaches to the water end use classification problem: (1) simple decision tree method based on three physical features of each event, namely volume, duration and flow-rate (e.g. *Trace Wizard* and *Identiflow*); (2) sensor networks on water end use appliances supported by data mining techniques (e.g. *Hydro Sense*); and (3) a hybrid combination of pattern recognition algorithms and data mining techniques to learn distinct flow signature patterns for each end use category to perform the classification process (e.g. *Autoflow*).

The first approach is resource intensive requiring significant analysis to disaggregate water end use patterns into discrete events accurately (Stewart et al. 2010). The second approach achieves high accuracy and does not require human interaction once the system is operating but requires sensors to be attached to many water use devices in the home, which makes this technique cost-intensive, intrusive (Froehlich et al. 2009, Nguyen et al. 2013b) and can artificially influence water use behaviour. The third approach (see Nguyen et al. 2013a, b; 2014) overcomes the deficiencies of the first two, by only requiring a smart meter installed at the property boundary. This approach uses pattern recognition (i.e. HMM and ANN) coupled with other data mining techniques (i.e. event probability analysis) to automate the end use analysis process. A software tool (i.e. Autoflow) was developed to provide a user-friendly platform to aid this process as described in the following paragraph.

The *Autoflow* approach has achieved average pattern recognition accuracy of 85%, which needs to be improved further prior to widespread application. While mechanised appliance (e.g. clothes washer) recognition accuracy is greater than 90%, behaviourally influenced end use categories such as shower, bathtub and irrigation, need to be improved to develop a more reliable system. This paper enhances previous research conducted by the researchers (Nguyen et al. 2013a, b; 2014) by combining HMM and Artificial Neural Network (ANN) to examine both the shape pattern and physical characteristics of each event in order to identify the most likely end use. The overall recognition process can be divided into three main stages: (1) likelihood estimation using HMM based on event flow-rate pattern; (2) examination of physical feature similarities (e.g. rate of change of water flow) using ANN; and (3) integration of HMM and ANN likelihood decisions to perform the final categorisation. Figure 1 illustrates a screen shot of the developed *Autoflow* software tool.

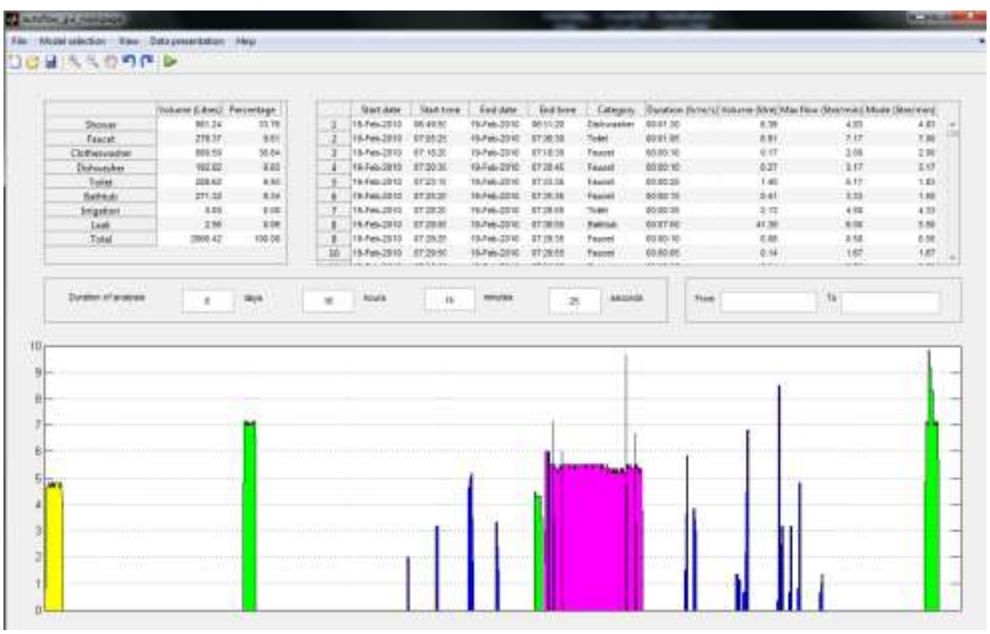
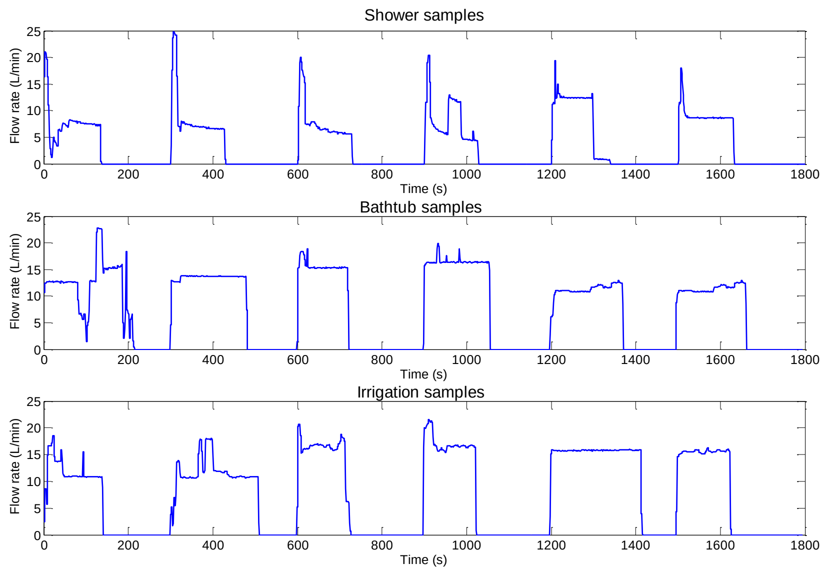


Figure 1. Autoflow (version 1) application illustrative example

2. Standalone application of HMM for water end use classification

A hidden Markov model (HMM) is a statistical Markov model in which the system is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered to be the simplest dynamic Bayesian network, which is one of the most popular techniques in the field of hand writing and speech recognition (Ephraim and Merhav, 2002). Principal theories and typical applications of this technique have been presented in Baum and Petrie (1966), Starner and Pentland (1995), Baum et al. (1970), Cho et al. (1995), Ghahramani and Jordan(1997), Chien and Wang (1997), Satish and Gururaj (2003) and Tapia (2004). In the development of *Autoflow*, HMM was applied as the main technique for pattern categorisation based on learning the distinct shapes of each end-use. Readers are referred to Nguyen et al. (2013a, b; 2014) for a full description of the application of the HMM technique for water end use classification.

However, with accuracy of only 85%, HMM applied alone will not be able to derive the levels of accuracy expected for a commercial software application. HMM does not adequately discriminate end use categories that are highly dependent on user behaviour (thus having similar features to other categories). To illustrate this issue, Figure 2 shows that three categories of end use event (i.e. shower, bath tub and irrigation) possess similar patterns despite



belonging to different categories. As a result, an additional technique that can inspect the physical features of these events is required to help discriminate between them.

Figure 2. Similar samples extracted from three different end-use categories

3. Event features used for decision making using ANN

ANN is one of the most powerful techniques in pattern recognition due to its adaptability to different types of problem models. ANNs are comprised of one or more processing units called 'artificial neurons' or 'perceptrons' (Karayiannis and Venetsanopoulos, 1993). Perceptrons of an

ANN are interconnected with one another by a series of weighted connections. The perceptrons of an ANN, depending on the system being replicated, are arranged in layers, with each perceptron of the preceding layer having a weighted connection with each neuron of the proceeding layer. In the process of ANN training to replicate a system, a training data set is fed through the network. Each perceptron processes the input data or input signal from either the input layer or the preceding perceptrons. The final layer of the ANN produces an output signal. The weights and structure of the network are altered in a manner depending on the specific training algorithm. In this study, a feed-forward network with back-propagation training algorithm is selected as the main tool to learn the typical pattern of each category.

To facilitate this task, nine different characteristics are proposed to describe each water event, including, (i) volume, (ii) duration, (iii) maximum flow-rate, (iv) most frequent flow-rate, (v) frequency of most frequent flow-rate, (vi) magnitude of initial flow-rate rise, (vii) magnitude of flow-rate drop at the end of event, (viii) gradient of initial flow-rate rise, and (ix) gradient of flow-rate drop at the end of event. Besides the first five features which have been employed in all existing flow-trace analysis applications, the inclusion of the last four can also be considered as critical features that can improve recognition accuracy. These features are defined or calculated as follows:

- *Magnitude of initial flow-rate rise (vi)* is defined as the flow-rate rise at the initial phase of the event when the consumer starts using water.
- *Magnitude of flow-rate drop at the end of event (vii)* is the flow-rate drop at the end phase of the event when the valve is shut.
- *Gradient of initial flow-rate rise (viii)* is determined by getting feature (vi) divided by the time it takes to reach this flow-rate.
- *Gradient of flow-rate drop at the end of event (ix)* is determined by getting feature (vii) divided by the time it takes for the flow to revert to zero.

4. Hybrid combination of HMM and ANN for water end-use classification

The database utilised in this study came from two recently completed end-use studies conducted in Brisbane (Beal and Stewart, 2011) and Melbourne (Gan and Redhead, 2013), Australia, where high resolution water meters recording 0.014 L/pulse at five second intervals was being recorded from over 500 homes. This data was manually disaggregated into end uses by analysts using the Trace Wizard™ software. The outcomes of this end use analysis process are seven different sets of main water patterns, including shower, faucet (tap), dishwasher, clothes washer, toilet, bathtub, and irrigation. The respective numbers of samples for each end use category are presented in Table 1.

To enable a classification process using ANN, the nine distinct features extracted from all samples will be used as the main input parameters for the ANN training process. The verification process on 16,000 samples is also presented in Table 1 and shows that once the model has been fully trained, it is able to assign a given independent sample into an appropriate end use category with an accuracy of at least 70%. This indicates that the ANN approach can categorise events with reasonable accuracy but alone is not adequate.

HMM has proven to be very suitable for classifying mechanised water end use categories such as clothes washer, dishwasher and toilet. However, lower levels of accuracy were accomplished when adapting these techniques on complicated shower, bathtub and irrigation events as their flow rate patterns and corresponding physical characteristics vary considerably.

Accurate water end use classification can be achieved if both the physical characteristics and shape features of a given sample can be integrated into a single intelligent model. The hybrid combination of HMM and ANN enables a much higher pattern recognition accuracy than standalone applications. Based on this rationale, a method that combines HMM and ANN was developed to help inspect the complete spectrum of water end use characteristics in order to make the most appropriate decision.

The final decision can be explained by the following logic. Given $\boldsymbol{A} = (a_1, a_2, \ldots, a_7)$ 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), or irrigation (a_7) using HMM, $\boldsymbol{B} = (b_1, b_2, \ldots, b_7)$ being the achieved likelihood when estimating this event with ANN, and $\boldsymbol{C} = (c_1, c_2 \ldots c_i, \ldots c_7)$ being the combined likelihood of A and B, where $c_i = a_i b_i$ and $(1 \leq i \leq 7)$, then the unclassified event will be assigned to category (i) if c_i is the maximum value of $\boldsymbol{C}$. This conclusion will be confirmed through a verification process presented in the next section.

5. Hybrid classification model accuracy verification

A model verification process was conducted that utilised 16,000 samples to illustrate the accuracy achieved for standalone HMM and ANN models as well as the hybrid HMM with ANN model. Table 1 demonstrates that a considerable increase in accuracy was achieved for the hybrid model. In terms of the more variable end use categories, significant improvements of 22.8%, 24.1% and 22.3% have been achieved for the shower, bathtub and irrigation end use categories, respectively. In addition, an increase of 14.6% was also obtained (76.2% with HMM compared to 90.8% using the combined model) for tap discrimination (another user-dependent category). As expected, the hybrid method achieved lower increases in accuracy for the clothes washer (8.1%), dishwasher (5.1%) and toilet (5.0%) end use categories as they produce relatively distinct patterns which can be correctly identified by HMM alone.

The developed hybrid method employed in the *Autoflow* software tool has been demonstrated to be more effective than existing methods that apply HMM as the sole classifier. Future studies will integrate various probability functions (e.g. event time of day probability) into this model to help further improve results.

Table 1. Model testing for Melbourne homes

Category	No. of samples for training	No. of samples for testing	Recognition accuracy (%)		
			HMM	ANN	HMM+ANN
1. Shower	14,903	3000	70.3	78.6	93.8
2. Faucet	21,985	3000	76.2	71.7	90.8
3. Clothes washer	15,211	3000	82.6	72.8	91.7
4. Dishwasher	13,342	3000	90.9	76.9	96.0
5. Toilet	15,222	3000	89.4	88.8	94.4
6. Bathtub	1,080	500	64.0	88.0	88.1
7. Irrigation	1,020	500	63.6	70.6	85.9

All categories	82,763	16,000	76.7	78.2	91.5

6. Vision of an intelligent water metering and information system

Present smart metering technologies allow water businesses to autonomously complete more sophisticated reporting of water consumption and disseminate this information near real-time to consumers. One of the major barriers to the widespread implementation of smart metering is the lack of novel hydroinformatic techniques to manipulate this abundant water use information into easily digestible forms. The final stage of this project aims to integrate these analysis modules into an intelligent system for residential water end-use classification, customer feedback and enhanced urban water management It is envisaged that the system could be interfaced with customers and water business managers via a web-portal or mobile phone applications.

The proposed system illustrated in Figure 3 will allow individual consumers to log into their user-defined water consumption web page to view their daily, weekly, and monthly consumption tables, as well as charts on their water demand across major end use categories (e.g. leaks, clothes washer, shower, irrigation, etc.). It could also rapidly alert customers to leaks rather than waiting for the present slow feedback process from the traditional metering technology (e.g. quarterly bill). This system would also benefit water businesses by rapidly providing water end-use reports at a range of scales. This may aid in the development of more targeted conservation programs, improved water demand forecasting and optimised pipe network modelling.

7. Conclusion

The combination of the ANN and HMM techniques for water end use classification has improved recognition accuracy levels over their individual application. The hybrid approach described will be codified and incorporated into the next version of the *Autoflow* software tool. This research will advance both the knowledge base and water industry practices and provide a significant contribution to the hydro informatics discipline, particularly in relation to novel pattern recognition algorithms. These algorithms will be applied to autonomously disaggregate domestic water flow data into end use events for both consumers and water businesses. These important outcomes will have far reaching implications on the way the water sector operates. Moreover, this research will contribute to the development of new urban water management systems that can deliver smart information to both customers and water businesses. This information will significantly enhance the relationship that customers have with their water consumption and contribute to the operational efficiency (and sustainability) of water businesses. It is likely that tools will be developed to provide customers with near real-time water use information (Figure 3). This could include analyses such as comparisons of shower use with other comparable households or being alerted to leaks and their likely cause (e.g. toilet seal). Water businesses may be able to better forecast water demand, evaluate appliance rebate programs (e.g. showerhead replacements) and better model and manage hydraulic networks.

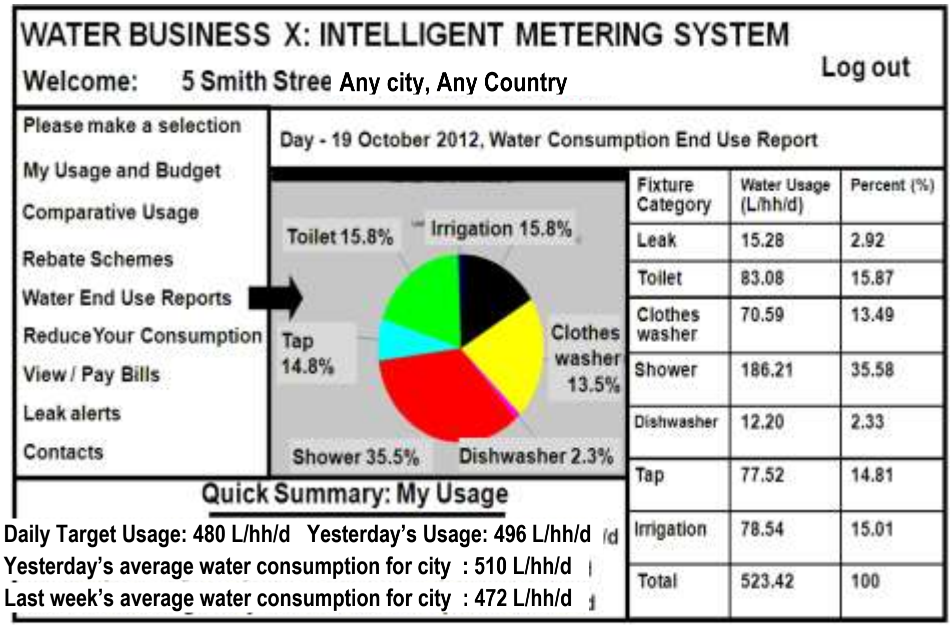


Figure 3. Illustrative interface for a proposed water management information system

Acknowledgements

Yarra Valley Water (Anthony Brown and Peter Roberts), South East Water (Lorraine Nelson and Gayathri Ramachandran) and City West Water (Nilmini Siriwardene and Chris Jones) are acknowledged for contributing data and funding to this research.

References

[1] Baum, L. E., Petrie, T, Soules, G and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1), 320-339

[2] Beal, C. and Stewart, R.A. (2011). South East Queensland residential end use study: final report. *Technical Report No. 47 for Urban Water Security Research Alliance*. Griffith University and Smart Water Research Centre, January 2012.

[3] Chien, J.-T., Wang, H.-C. (1997). Telephone speech recognition based on Bayesian adaptation of hidden Markov models. *Speech Communication*, 22, 369-384.

[4] Cho, W., Lee, S.W., and Kim, J.H. (1995). Modelling and recognition of cursive words with HMM. *Pattern Recognition*, 28(12), 1941-1953.

[5] Ephraim, Y., Merhav, N. (2002). Hidden Markov processes. *Information Theory*, IEEE Transactions on, 48, 1518-1569.

[6] Froehlich, J.E., Larson, E. (2009). HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. In: Proc. of UbiComp 2009, Orlando, Florida, USA, 235–244 .

[7] Gan, M. and Redhead, M. (2013). Melbourne Residential Water Use Studies. Smart Water Fund, June 2013, 10TR5-001.

[8] Ghahramani, Z., Jordan, M. I. (1997). Factorial Hidden Markov Models. *Machine Learning* 29 (2/3): 245–273

[9] Grimm,N. B., Faeth,S. H., Golubiewski,N. E., Redman,C. L., Wu,J., Bai,X., Briggs,J. M. (2008). Global change and the ecology of cities. *Science*, 319: 756-760

[10] Nguyen, K.A., Stewart, R.A. and Zhang H. (2013a). Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-Environment Research*, 7(3), 182-2001.

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

[12] Nguyen, K.A., Stewart, R.A. and Zhang H. (2014). An autonomous and intelligent expert system for residential water end-use classification. *Journal of Expert Systems with Application*, 41(2), 342-356.

[13] Satish, L. and Gururaj, B. I. (2003). Use of hidden Markov models for partial discharge pattern classification. *IEEE Transactions on Dielectrics and Electrical Insulation*, 28(2), 172-182.

[14] Starner, T., Pentland, A. (1995). Real-Time American Sign Language Visual Recognition From Video Using Hidden Markov Models. *Master's Thesis*, MIT, Program in Media Arts.

[15] Stewart, R.A., Willis, R.M., Giurco, D., Panuwatwanich, K., and Capati, B. (2010). Web-based knowledge management system: linking smart metering to the future of urban water planning. *Australian Planner*, 47(2), 66-74.

[16] Tapia, E., Intille, S.S., and `Larson, K. (2004). Activity Recognition in the Home Using Simple and Ubiquitous Sensors. In: Ferscha, A., Mattern, F. (eds.) PERVASIVE 2004. LNCS, vol. 3001, pp. 158–175. Springer, Heidelberg.