Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 119 (2015) 1391 – 1398

Procedia Engineering

www.elsevier.com/locate/procedia

13th Computer Control for Water Industry Conference, CCWI 2015

A Benchmarking Model for Household Water Consumption Based on Adaptive Logic Networks

Xiaomin Chen[a], Shuang-Hua Yang[a,*], Lili Yang[b], Xi Chen[a]

[a]Department of Computer Science, Loughborough University, LE11 3TU, UK
[b]School of Business and Economics, Loughborough University, LE11 3TU, UK

Abstract

Household water benchmarking is an important step in evaluating a household's water usage and comparing it with similar households. It can provide an indicator if a household consumes more water than usual during a certain period of time or some households consume more than other similar households in a particular region. This paper proposes a benchmarking model for household water consumption based on Adaptive Logic Networks (ALNs). Real world data collected by a water consumption monitoring system installed in Sosnowiec, Poland and Skiathos, Greece is respectively used to build a model for each city. The results indicate that the developed models can successfully prediction for a particular use purpose.

Peer-review under responsibility of the Scientific Committee of CCWI 2015

Keywords: Household water consumption; Benchmarking; Socio-demographical factors; Adaptive Logic Networks

1. Introduction

Water has been identified as one of the most significant natural resources and a key to prosperity and wealth. As Marshall [1] remarked, water has played a crucial role in the location, function and growth of communities. However the World Economic Forum has announced in 2015 that water crisis ranks the eighth global risk with the highest likelihood of occurring within 10 years [2]. In the light of these facts, any cause of wasting water should be identified and removed if possible. During daily routines, water usage at a household level can be effectively reduced by increasing consumers' awareness and changing their inappropriate habits, e.g. brushing teeth with the tap running, using toilet as a dustbin, leaving a leaky faucet unfixed and etc. To cut off unnecessary water wastage, a benchmarking model is required to provide a baseline measure with which the water consumption can be compared. Based on the comparison result, water wastage can be identified, and accordingly the intervention strategy can be designed and deployed.

There has been a great deal of research conducted on water benchmarking. [3] gives a comprehensive review on the state-of-the-art water benchmarking methodologies. In this paper we propose a benchmarking model based on Adaptive Logic Networks (ALNs). It takes the socio-demographical information as inputs and outputs a prediction

* Corresponding author. Tel.: +44 (0) 1509 635670.
E-mail address: S.H.Yang@lboro.ac.uk

Peer-review under responsibility of the Scientific Committee of CCWI 2015
doi:10.1016/j.proeng.2015.08.998

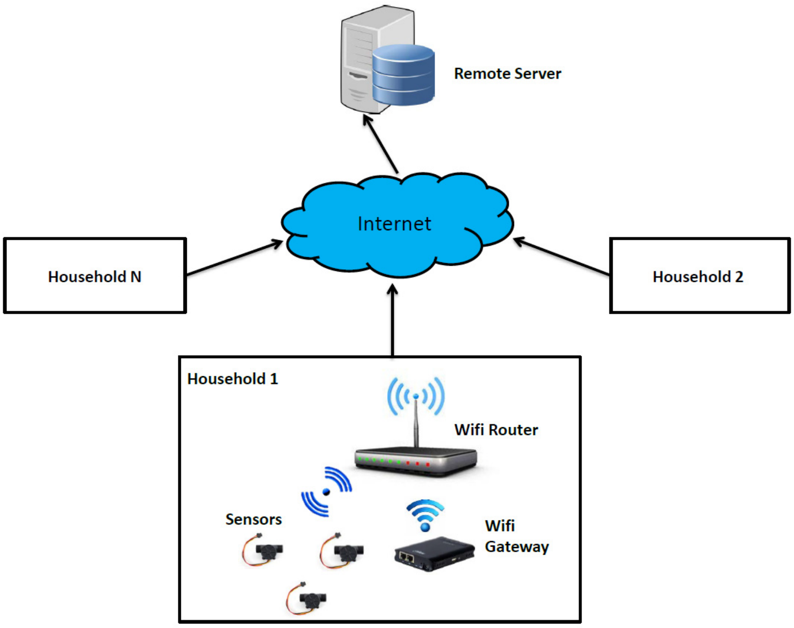

Fig. 1. Water consumption monitoring system

on the average water usage. Different from most of the benchmarking tools which are built using statistical data collected from surveys, in this work we use real world water consumption data collected by a monitoring system. It can provide a baseline for comparison within a single household over time or across multiple households in a region, and help water management organisations to diagnose unusual water usage, target potential saving opportunities and assess their water resource management and development policies. The historic water consumption data is used to establish a benchmarking model for households in a region. To evaluate the water consumption for a household whose historic data has been used to build a model, the new consumption is compared with the model prediction, and water wastage can then be identified. It is actually a comparison over time. Besides, the model can be used for comparison across similar households in the same region. If the data of a household has not been used for building the model, the consumption can be evaluated by entering its corresponding input values into the model. The model outputs a prediction based on the historic data of households with similar input values. The output provides a reference on the amount of water this household is supposed to use. If the actual usage is above the reference, it indicates that there might be some water wastage.

2. Data collection

A household water consumption monitoring system is designed for collecting detailed information on the amount and the way water is used in a household. The system, as shown in Fig. 1, consists of a local wireless monitoring system and a remote central server. The local wireless monitoring system includes a few sensors, a wifi router and a wifi gateway. Sensors are installed on water supply pipes to detect flow rates at different spots in a house, such as kitchen tap, dishwasher, washing machine, shower, toilet and etc.. The detected data is transmitted to the remote central server through the wifi gateway and the wifi router. The remote central server receives data from multiple household monitoring systems. The collected data is stored in a database. One can access the data via logging on to the database.

The developed monitoring systems were installed in 30 recruited households, 10 in Sosnowiec, Poland and 20 in Skiathos, Greece. Near real-time monitoring of water consumption in the 30 households has been carried out from

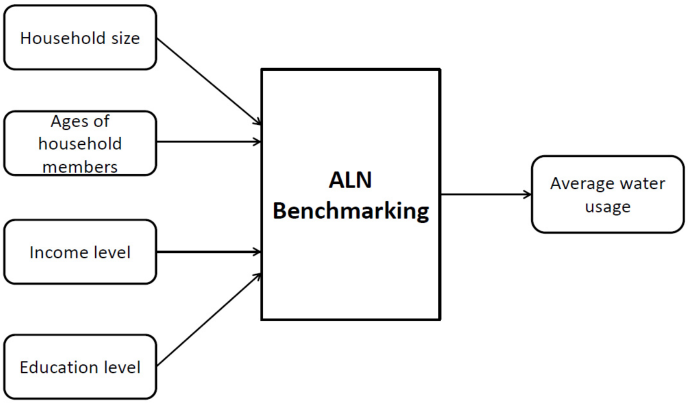

Fig. 2. ALN based benchmarking model

March 2015 on. A great deal of data has been inserted to the database, which can be used for deriving the water consumption pattern and further benchmarking modelling.

3. Factors influencing household water consumption

The household water consumption is determined by quite a few factors, such as climate, seasonality, socio-economic characteristics, socio-demographics, etc.. In this work we only take the socio-demographic factors into account, because the data available for modelling only lasts two months, during which there were no significant climate or seasonal fluctuations, and the water price remained the same. Socio-demographic information is expected to have the dominant impact on the water consumption. The purpose of a benchmarking model is to relate the socio-demographic factors to the household water consumption. As shown in Fig. 2 the inputs are the influencing factors, and the output is the corresponding household average water consumption. Given a set of input values, the benchmarking model will be able to provide a prediction on the corresponding water consumption value.

Four socio-demographical factors are selected as determinants to build the model, which are respectively the household size, ages of household members, income level and education level. The socio-demographic information for the 30 recruited households was collected via interviewing.

3.1. Household size

The number of household members affects the amount of water used in a house [4,5]. A household of a large size normally uses more appliances with greater frequency, resulting in more water usage than a small size household. [6] found that the water consumption increases with the household size, although it is not a proportional increase. For example, a household with two people use less water than a household with four people, but not half as less. However, household size was found to be an insignificant factor of water usage in [7]. This finding does not necessarily contradict findings of other researchers, but rather reflects the fact that interior water use is substantially less than outdoor uses. Household size then may not affect exterior water use, e.g. a house with only one resident may have as much lawn to water as a household with four residents. Similarly, [8] found that increasing household size was associated with improved interior water efficiency, resulting in per capita savings despite overall household consumption increasing.

3.2. Ages of household members

The water use behaviours can be quite different among different ages of household members. Households with children could be expected to use more water. Youngsters might use water less carefully, e.g. taking more showers, doing more frequent laundering, while retired people might be much thriftier. Theses expectations are confirmed by studies such as [9]. We use the average age of household members as an age indicator factor in the modelling.

3.3. Income level

Income is a significant factor in water consumption. [7] indicates that income rises result in a corresponding increase in water consumption, but again not a proportional increase [10,11]. Wealthier households are likely to have more water consuming appliances, swimming pools, and larger lots. Richer consumers tend to have a lower level of perception of the rate structure, as the total bill represents a lower proportion of their income. But significant differences in personal water use habits in households with different incomes have not been found. In this work we use the household total income as an income-related input.

3.4. Education level

The education level influences the water consumption in a household as well. It has been shown in [12] that the education level is positively correlated with lower water consumption and higher water conservation behaviours. Educated people tend to be more aware of the importance of water resource, more informed of the global water crisis, and thus more eco-conscious. Highly educated parents might pass the right water use habits to their children, which would further cut down the household total water consumption. The education levels from primary school to PhD are respectively marked from 1 to 5 in this work. The input to the model is the average education level of adults above 18 years old. Children under 18 are not included.

4. Adaptive logic networks modelling

4.1. Adaptive logic networks

Adaptive Logic Networks (ALNs) were proposed by Armstrong in [13]. An ALN is a multi-layer perceptron (MLP) feedforward network exactly like a Neural Network but with a few simplifications. The nodes have two input leads, and the input signals are Boolean. Each connection weight is determined by a single bit of information and the 'squashing function' is a threshold [14]. An ALN is actually made up of trees with nested MIN and MAX operators on multivariate hyper-planes. It is an analytical tool which expands and combines the power of Neural Networks with classical statistics.

ALNs have three distinct advantages over Neural Networks as they are far more computationally efficient, offer easily interpreted results and allow the users to examine the properties of the network and vary them. From a statistical viewpoint an ALN is a good model to use as it allows statistically unimportant inputs to be removed and furthermore restrictions can be put on the model before it is even formed.

ALNs can be used to model any complex, nonlinear relationship between a dependent variable and one or more independent variables, and is computationally efficient. The household water consumption and the influencing factors fit this relationship, and therefore we choose an ALN as a tool to build the benchmarking model.

4.2. Network architecture

Network architecture determines the number of connections in a network and how information flows through the network. It has a large impact upon the accuracy of the network output. In this work we use only one hidden layer. It has been proven that ALNs with one hidden boundary layer can approximate any problem given sufficient connection weights and nodes. The number of nodes in the input layer is fixed by the number of input factors, and the number of output is one, i.e. the household water consumption. The number of nodes in the hidden layer and the connection weights are determined in the process of training and validation.

4.3. Modelling procedure

The steps to build an ALN model are described as follows:

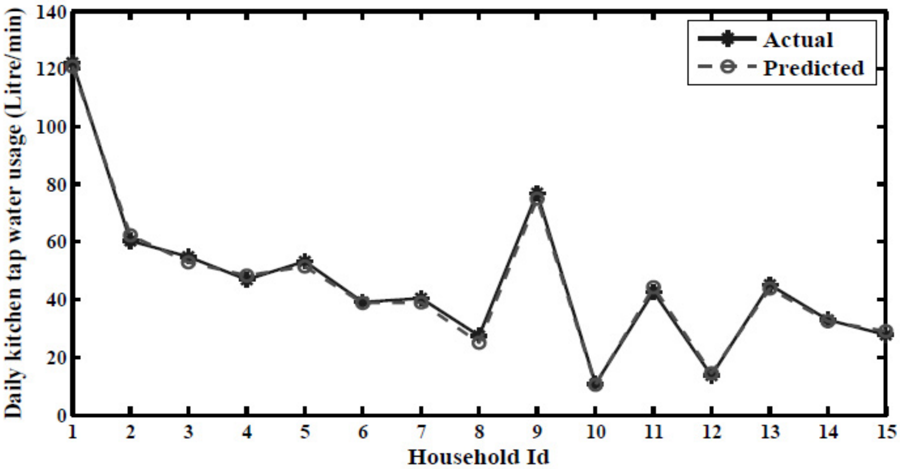

Fig. 3. Comparison of ALN predicted daily kitchen tap usage to the actual value for Greece sample households

1. Take 10% out of the raw data set for testing the accuracy of the final ALN;
2. Divide the remainder of the raw data into two sets, 70% for training and 30% for validation;
3. Fit a multiple linear regression using the training set [15];
4. Split one piece into two pieces if the RMS error above error tolerance;
5. Check with validation set, remove split if over-fitting, otherwise go back to step 3;
6. Repeat step 3 - 5 to produce multiple ALNs (10 ALNs in this work);
7. Average out the multiple ALNs to produce the best one, known as "bagging".

5. Results and performance evaluation

Using the real world data collected by the water consumption monitoring system mentioned in Section 2 and the socio-demographic information collected through interviewing, an ALN benchmarking model for a particular water use purpose can be established for sample households in each city.

The kitchen tap water data of Skiathos, Greece from 24/04/2015 to 12/05/2015 is used to build an ALN benchmarking model to evaluate the kitchen tap usage for each household. The output of the ALN model is the conditional expectation of the dependent variable given the independent variables. In this case, it is the mean daily kitchen tap water usage given a set of socio-demographic input values of a household. Fig. 3 plots the predicted values from the ALN model and the actual values for the testing data set.

The model performance is evaluated using the percentage error $PE_i = |y_i - \hat{y}_i|/y_i$ for Household i, in which $\hat{y}_i$ is the predicted value, and y_i is the actual value. The PE values for the 15 households are listed in Table 1. It can be seen that all of the PE values are below 10%, and over 85% of them are below 5%. This indicates that the developed ALN model gives a highly accurate prediction.

Table 1. Percentage errors of 15 households for the Greece kitchen tap usage model

Household ID	1	2	3	4	5	6	7	8
PE	0.0117	0.0330	0.0369	0.0305	0.0342	0.0064	0.0360	0.0874
Household ID	9	10	11	12	13	14	15	
PE	0.0268	0.0304	0.0391	0.0572	0.0302	0.0132	0.0467	

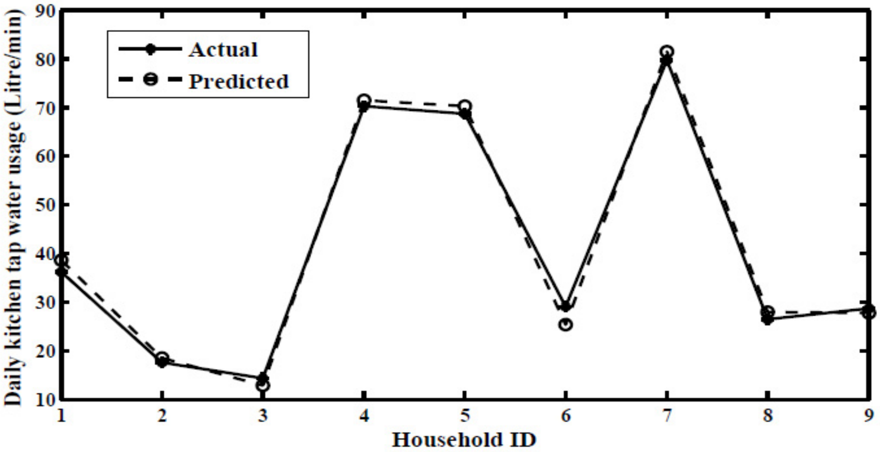

Fig. 4. Comparison of ALN predicted daily kitchen tap usage to the actual value for Poland sample households

Another ALN model is built for the kitchen tap water usage in Sosnowiec, Poland. The data is collected from 06/03/2015 to 19/05/2015. The predicted values and the actual values for the testing data are plotted in Fig. 4. The corresponding PE values are listed in Table 2, out of which 2 are above 10%, 3 are between 5% and 10%, and 4 are below 5%. Although the prediction errors of this model are slightly larger than those of the model for Greece kitchen tap water, it can still be viewed as a good model with acceptable error ranges.

Table 2. Percentage errors of 9 households for the Poland kitchen tap usage model

Household ID	1	2	3	4	5	6	7	8	9
PE	0.0665	0.0536	0.1045	0.0174	0.0225	0.1263	0.0216	0.0560	0.0308

The dishwasher data from 06/03/2015 to 19/05/2015 of Sosnowiec, Poland is also used to build a benchmarking model. Fig. 5 plots the predicted and the actual values, and the corresponding PE values are listed in Table. 3. It can be seen that the performance of this model is quite good with only one PE above 10%.

Table 3. Percentage errors of 9 households for the Poland dishwasher water usage model

Household ID	1	2	3	4	5	6	7	8
PE	0.1055	0.0126	0.0313	0.0241	0.0141	0.0026	0.0867	0.0259

It is noted that the number of households whose data is used to build the above models is less than that of recruited households in the monitoring system because the data set for some households is not large enough to build an accurate prediction model due to unexpected network disconnections during the period data was collected.

6. Discuss

The output of the ALN benchmarking model can serve as a baseline measure for intervention strategy design. The historic data is used to build a model for households in a particular region. To evaluate the water consumption behaviours, the socio-demographic information of a household is input to the model. A prediction on the mean water consumption value is generated by the model. The predicted mean can then be compared with the actual mean

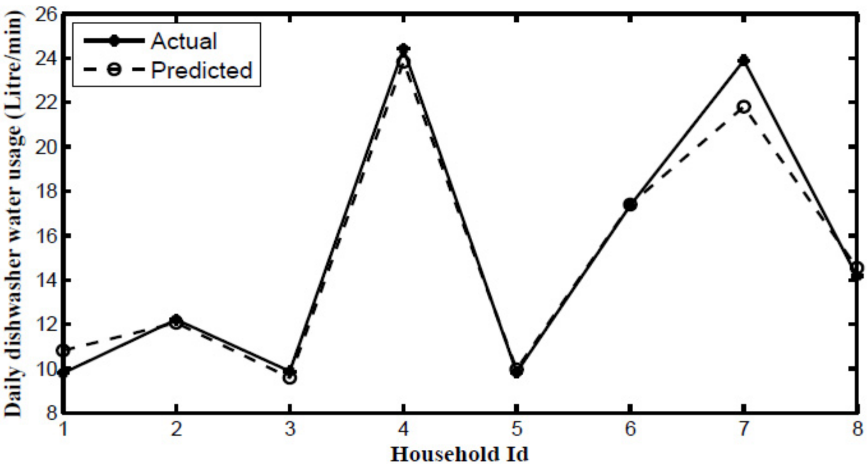

Fig. 5. Comparison of ALN predicted daily dishwasher water usage to the actual value for Poland sample households

calculated based on meter readings for a certain period of time. If the actual mean is larger than the prediction, it gives an indication that for that period of time, this household consumes more water than usual. Further analysis and investigation need to be carried out to find the reasons, and the intervention strategies need to be deployed accordingly.

Section 5 gives a description on the modelling methodology, and provides results based on limited amount of real world data. The model can be easily expanded to include more households, e.g. hundreds, with provision of richer data set. The model will be more generic and provide more accurate predictions.

Besides the single water use purpose as discussed in Section 5, a benchmarking model for the total household water consumption can be built just by using the total consumption historic data for training. Not only the total consumption, the output can also be the sum consumption of a few purposes, e.g. the kitchen water usage including both the kitchen tap and the dishwasher, as long as the specific historic data is chosen for training.

The model described in this paper only takes the socio-demographic information as inputs. It actually averages out some uncertainties, e.g. the seasonal changes, the week of the day and etc. To improve the prediction accuracy, the model can be further refined in terms of different concerns. For instance, to account for the seasonal changes, the model needs to be updated on a regular basis. The simplest way is to update twice a year, one for hot season and the other for cold season. Each time the model is trained using the data in season. The output is thus the average water usage for the current season, which is more accurate than the average over the whole year.

7. Conclusions

This paper proposes a benchmarking model based on Adaptive Logic Networks. It can provide a baseline for comparison within a single household over time or across multiple households in the same region. The model takes account of the socio-demographical information as input and outputs a prediction on the average household water usage. Real world water consumption data collected in Sosnowiec, Poland and Skiathos, Greece by a monitoring system is used to build the model. The results indicate that the developed models can successfully predict the water usage for a particular use purpose in the two cities.

Acknowledgements

This work is part of the ISS-EWATUS project (issewatus.eu) and has been funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no (619228).

References

[1] A. Marshall, Water as an element of national wealth, in Pigou, A. (Ed.), Memorials of Alfred Marshall, Kelley and Millman, New York, reprinted in 1956, pp. 134-141.
[2] World Economic Forum (2015).Global Risks 2015 Report. Jan, 2015.
[3] S. Berg, Water Utility Benchmarking: Measurement, Methodologies, and Performance Incentives, IWA Publishing, 2010.
[4] S. Gaudin, Effect of price information on residential water demand, Applied economics, 38(4), 2006, pp. 383-393.
[5] E. Wentz and P. Gober, Determinants of small-area water consumption for the city of Phoenix, Arizona. Water Resources Management, 21(11), 2007, pp. 1849-1863.
[6] F. Arbus, R. Barbern and I. Villana, Price impact on urban residential water demand: A dynamic panel data approach, Water Resources Research, 40(11), 2004.
[7] S. Guhathakurta and P. Gober, The impact of the Phoenix urban heat island on residential water use, Journal of the American Planning Association, 73(3), 2007, pp. 317-329.
[8] A. S. Polebitski and R. N. Palmer, Seasonal residential water demand forecasting for census tracts, Journal of water resources planning and management, 136(1), 2010, 2736.
[9] C. Nauges and A. Thomas, Privately operated water utilities, municipal price negotiation, and estimation of residential water demand: the case of France, Land Economics, 76(1), 2000, pp. 68-85.
[10] J. M. Dalhuisen, R. J. G. M. Florax, H. L. F. de Groot and P. Nijkamp, Price and income elasticities of residential water demand: a meta-analysis. Land Economics, 79(2), 2003, 292308.
[11] M. E. Renwick and R. D.Green, Do residential water demand side management policies measure up? an analysis of eight california water agencies, Journal of Environmental Economics and Management, Elsevier, 40(1), 2000, pp 37-55.
[12] K. Millock and C. Nauges, Household adoption of water-efficient equipment: the role of socio-economic factors, environmental attitudes and policy. Environmental Resource Economics, 46, 2010, pp. 539-565.
[13] W. Armstrong and M. Thomas, Adaptive logic networks, The Handbook of Neural Computation,Section C1.8, 1997.
[14] W. Armstrong, A. Dwelly, J. Liang, O. Liang, D. Lin and S. Reynolds, Some results concerning adaptive logic netowrks, Internal report Department of Computing Science, The University of Alberta, 1991.
[15] K. Cogger, Nonlinear multiple regression methods: a survey and extensions, Intelligent Systems in Accounting, Finance and Management, 17, pp.19-39, 2010.