Pattern Recognition in Residential End Uses of Water Using Artificial Neural Networks and Other Machine-Learning Techniques

Juan Carlos Ibáñez Carranza[1], Roberto Díaz Morales[2], Jose Antonio Sánchez[3]
[1] Canal de Isabel II. Santa Engracia, 125. 28003 Madrid (Spain)
[2] Treelogic. Parque Tecnológico de Asturias, parcela 30, 33428 Llanera, Asturias (Spain)
[1]jci@canaldeisabelsegunda.es [2]roberto.diaz@treelogic.com [3]joseantonio.sanchez@treelogic.com

ABSTRACT

Machine-Learning and other Artificial Intelligence techniques have nowadays many practical applications in engineering, science or everyday life. In the water industry, there is also a broad scope of potential applications. In this paper, it will be presented a system developed by Canal de Isabel II to identify residential use of water in its different appliances, based on records from precision water meters equipped with pulse emitters. Developed models are based on Support Vector Machines, and Artificial Neural Network paradigms. Training data sets for the models have been extracted from a sample of about 300 residential users in the Region of Madrid (Spain), monitored since 2008. In this time, more than 35 million of water use events have been registered and about 15 million hours of water consumption monitored.

Machine-Learning techniques have proved to be an accurate and suitable method for automate this task that otherwise should require a huge number of man-hours of processing by operators.

Keywords: Water End Use; Residential Water Use; Artificial Neural Networks; Support Vector Machines; Machine Learning.

1 BACKGROUND

Machine learning [4] is a type of artificial intelligence (AI) that provides computers with the ability to solve a task learn without being explicitly programmed. Machine Learning techniques are currently the state of art to solve many problems in a wide variety of research fields (e.g. speech and handwriting recognition, fraud detection, marketing, DNA sequence determination or self-driving cars). In the water industry, although used to more traditional methods for research, there is a broad scope of potential applications of advanced methods of artificial intelligence.

Indeed, there are some previous works published for this problem. In [5], a hybrid combination of techniques including Hidden Markov Model (HMM) and Dynamic Time Warping (DTW) to classify water end use is presented. In [6], the same authors include Artificial Neural Networks (ANNs) to improve their previous results.

Since ANNs have obtained some previous good results and due that Deep Learning (DL) is the state-of-art paradigm of Artificial Neural Network (ANN) computing, DL was our many technique to explore. In addition, as a benchmark, we implemented a similar approach but using Support Vector Machines (SVMs) as classifiers.

In this paper, it will be presented a system developed by Canal de Isabel II, aimed to identify end uses of residential water from the signals transmitted by precision water meters equipped with digital pulse emisors.

1.1 Water Demand Assessment and Management

Nowadays there is no doubt that sustainable management of the urban water supply should be based on demand management policies. The increase in the supply from new water sources will always represent more expenses, poorer quality and more environmental damage. It has been said [2] that "*the largest, least expensive, and most environmentally sound source of water* [...] *is the water currently being wasted in every sector of our economy*".

Efficient management of drinking water supply systems and design of demand management policies for a sustainable use of water requires of a good understanding of demand, its characterization, patterns of behavior and explanatory factors.

Design of demand management policies require answers to questions such as: What is the true potential for water conservation and efficiency improvements? How much water can we save in each economical sector? How much water can be saved by replacing actual domestic appliances with more efficient ones? These responses can only be found with a deep and detailed investigation about the current uses of water in the urban areas.

1.2 Residential End Uses of Water

Residential use of water represents about 70% of total consumption in an urban water supply system. Identify the amount of water used in taps, toilets, showers, clothes washer, dishwasher, irrigation, etc., their frequency and duration of use, habits of users and relation to factors such as weather and others can be of a great interest to water utilities, mainly in which relates demand management, and forecast for the short and long term.

Since it is not possible to obtain such information from the whole of the company's customers, due to technical limitations, a good approach can be reached from a representative sample.

Several methods can be applied: surveys on a sample of users is a non-expensive method although not very accurate. Direct measures in dwellings is a more accurate and scientific approach. However, installing a meter at each water using device in a home can be a highly intrusive technique and in practice unfeasible. Preferred methodology should consist in the installation of a unique precision meter at the dwelling main water connection, which records flow in a real or near real time basis. Specialised software should be used in order to process that information and deduce which device or devices are being used in every moment.

Canal de Isabel II [3] maintains since 2008 a sample of users consisting of about 300 dwellings spread over all the region of Madrid in Spain, which are being continuously monitored in order to determine patterns of consumption and end uses of drinking water supplied. This sample is sufficiently stratified and spread by the different geographical areas of the region to be considered representative of the domestic users of the Community of Madrid.

All dwellings of the sample are equipped with precision volumetric meters and electronic pulse transmiter which registers 1 or 10 pulses per litre. From this panel of users, they have been collected data of more than 35 million of events[1] of water use in about 15 million hours monitored for 9 years.

Software commercially available for processing this kind of data, such as *Trace Wizard®* (Aquacraft), or *Identiflow* are based on simple decision trees to classify water use events, and

[1] An event is a differentiated use of a certain water using appliance, and can be continuous, such as the recharge of a toilet tank, or discontinuous, like the use of a clothes washer involving several wash and rinse cycles.

require the supervision of an analyst, which involves a considerable amount of man-hours of process: about 1 hour of an operator to analyse a two-week period of data from each installation [4], which makes it unfeasible to deal with huge amounts of data when several hundreds of dwellings are concerned.

In 2009, Canal de Isabel II developed an automatic system to identify end uses of water, based on bayesian networks and minimal entropy heuristic for discretization of the samples data series. Periods of at least 2 – 3 months for each dwelling, previously processed by operators were used as training – test data sets. These data sets are considered the “ground truth” for validation of the bayesian models.

The system, herein presented, was developed in 2016, and was intended as an improvement over the previous design, using more advanced Machine Learning techniques.

2 METHODS

Two different techniques, based on supervised learning models have been investigated for classification of water end uses: Support Vector Machines (SVM) and Artificial Neural Networks (ANN). Both methodologies will be presented in detail in this paper.

Previously, it was necessary to develop original mathematical algorithms for transformation of the pulses emitted by the meters in instant flow, and to isolate and typify water use events.

Meters used to record water consumption were of volumetric rotary-piston type, equipped with pulse emitters that produce a pulse every time a certain volume of water (1 liter or 0.1 liter) passes through the device. Calculation of instant flow from volumetric pulses is not a trivial task, because of the uncertainty in determination of the precise instant when flow begins, till the volume completes the precision of the pulse emitter (1, or 0.1 liter). A mathematical algorithm was developed *ad hoc* to solve this question.

Once a continuous flow-time line is calculated, it is necessary to identify and isolate the different events of water use, considering that sometimes different water uses may be overlapped, for example, the recharge of a toilet tank can coincide in time with the use of taps or washing machines, etc. Also, an original algorithm was developed for this task.

Classification of the different events of water use, according to the device being used (taps, toilets, showers, clothes washer, irrigation, leaks, etc.) will be made based on the characteristics of the event. A total of 37 variables have been identified that could be used to characterize the events. They consider, duration, average and maximum flow, parameters that define the shape of the event, ascending and descending gradients and duration, and also other parameters related to previous and simultaneous events.

2.1 Support Vector Machines (SVM)

Kernel Methods [7] are a very popular family of Machine Learning algorithms. The have been widely used due to their ability to easily adapt linear models to create non-linear solutions by using the well-known ’kernel trick’, i.e. transforming the input data space onto a high dimensional one where inner products between projected vectors can be computed using a kernel function. Kernel methods have proved their effectiveness by obtaining highly competitive results in many practical problems.

Among them, SVMs are algorithms for supervised learning commonly used in classification tasks. The main idea behind a SVM is to create a hyperplane that separates two different classes of data while maximizing the margin (distance from the separating hyperplane to the closest pattern of every class). Patterns that do not respect this margin distance or directly are wrongly classified are called Support Vectors (SVs).

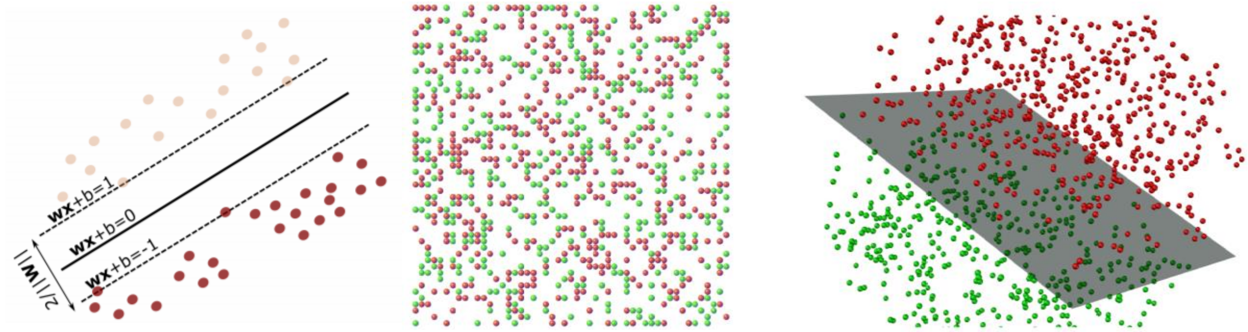

Figure 1. Maximum hyperplane separator, Non linear separable problem and a higher dimensional space where linear separation is possible

In real world problems, we find that most of them are not linearly separable. If we relax the formulation and we add a hinge loss function [8] the SVM separates the training data while some samples are still inside the margin or in the wrong side of the hyperplane.

Applying the ’kernel trick’, we can obtain non-linear solutions very easily by finding a higher dimensional space. To do that, in this project we have used a Gaussian kernel.

Figure 1 shows an illustrative example of the maximum margin classifier, a non-linear classification and a representation on a higher dimensional space.

As previously mentioned, the SVMs are binary classifiers. It is possible to adapt a binary classification method for multiclass problems. Techniques like **one-against-all** or **one-against-all** can be used. One-against-one is the preferred strategy in SVMs, as the training time is commonly shorter than one-against-all. The support vector machine training time depends on the number of elements in the training set. Using **one-against-one**, more models are to be trained but using fewer data than with **one-against-all** method.

2.2 Artificial Neural Networks (ANN) with deep learning

An ANN is a mathematical tool that, in a simplified manner models the human brain’s operation. In simplistic terms, it is a series of mathematical operations on an input vector that results in another output vector. The basic computation element is called neuron and although there are many types of neurons the most widespread one is known as perceptron.

A perceptron is only capable to solve data problems that are linearly separable in the input variables space. In order to avoid this limitation, an ANN is formed by several layers of neurons. When a non-linear operation is performed (e.g. sigmoid, rectifier linear unit, …) in the neuron, the ANN can solve non-linear problems.

Backpropagation is the method to calculate the gradient of the loss function with respect to the weights in an artificial neural network. Using this gradient we can use a gradient descent algorithm to iteratively train the network to solve a problem. One problem with the backpropagation algorithms is that the error is vanished exponentially as it crosses layers on its way to the beginning of the network. This is a problem because in a very deep network (with a lot of hidden layers), only the last layers are trained, whereas the first remained virtually unchanged.

Deep Learning (DL) is an area focused on the training of ANN with many layers. It has been recognized as one of ten breakthrough technologies according to MIT Technology Review [9]. DL attempts to model high-level abstractions in data with multiple non-linear transformations.

Stacked Denoising Autoencoders (SDAs)

SDAs [10] are a strategy to speed up the training procedure of deep networks. A **denoising autoencoder** (DA) is a neuronal network with a single hidden layer that learns to produce exactly the same information on the output as it receives on the input (where we have previously added noise). This leads the algorithm to discover important relationships that remain similar across all data.

A DAs finds relationships among the dataset features. For that reason, the weights inside the DA provide relevant information about the nature of the dataset. Using these weights to initialize a neural network by stacking several DAs speeds up the training procedure of deep ANN because it has already learnt to detect useful information (because ANN commonly initialize their weights using random values before the training).

Summarizing, the idea of *Deep Learning* via stacked autoencoders is precisely that: using several DAs and training them one by one, using each trained coder to initialize every layer and to train the next DA. After that, the training of a deep ANN is faster than using a random initialization.

The Final Model

To speed up the training procedure event more, we also included the output of some faster algorithms (GBT) as new input features of our model. Figure 2 shows the final architecture of our model:

- 37 input neurons. One for each input variable.
- 20 neurons in the first hidden layer.
- 100 neurons in the second hidden layer.
- 20 neurons in the third layer + 8 input neurons of the gradient boosted trees.
- 8 neurons in the output layer. One for each kind of event.

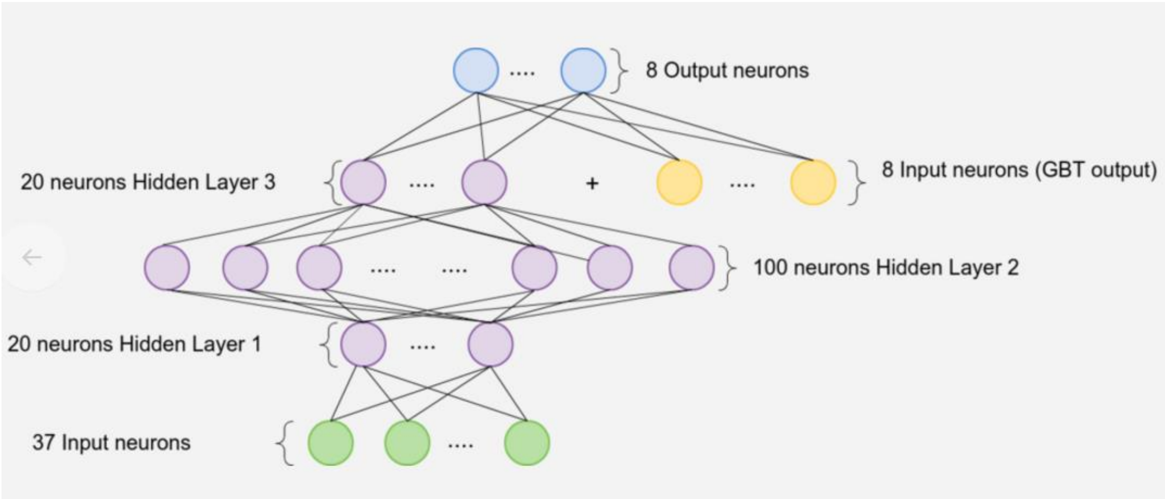

Figure 2. Artificial Neural Networks. The final model.

3 RESULTS AND CONCLUSIONS

The algorithm based on SDAs has been tested against SVMs. These experimental results show that it outperforms SVMs in this task. This section summarizes and compares the results obtained with both methods for the different types of water meters (1 and 0.1 liter).

Validation and test criteria

To correctly validate the performance of the algorithms, a validation and testing process was carried out. The data set used for creating and validating the models were the 2 – 3 months periods for each dwelling, previously processed by operators. These data sets are considered the “ground truth” for training and testing the SVM models and cover more than 600,000 hours of water consumption of 355 users.

This set of data was divided into two parts: training set (70% of the observations) and test set (the remaining 30%).

Validation and Training

The aim of the first phase is to obtain the value of every algorithm hyperparameters (the cost and the kernel parameter in SVM, the layers and learning rate with SDAs). To calculate the best value for these parameters, 75% of the training data were used in a previous training, and the remaining 25% to validate the result. The parameters achieving the best validation hit rate are those selected. After obtaining the values of these parameters, final training is performed with all the data.

Due to the computational cost of the SVM, with a complexity of $O(n^3)$ and the large number of meters to process, it was considered to limit the maximum training data value. This value has been set by default at 10,000 events. There are alternatives to improve their scalability [11] but they require expensive HPC hardware.

Testing phase

A test set is required to evaluate the performance of the method in events that have not previously been used for training.

The final results of the performance of the algorithm are calculated by evaluation on this test set. They are summarized in Table 1.

Table 1. Precision achieved in comparison with the validation data sets.

	SVM		**SDA**	
Meter type	N. events	liters	N.events	liters
1 liter	67.4%	63.4%	81.8%	85.8%
0.1 liter	84.8%	73.5%	91.2%	85.9%

Meters with 1 pulse/liter precision

The study has been done on 239 counters of 1 pulse/liter water meters. The results obtained with the SDA models are significantly better than those obtained with SVM. In summary, an average accuracy of 67.41% was obtained with SVM, while 81.78% was obtained with SDAs. This accuracy has been calculated taking into account the number of correctly classified events. The average accuracy in terms of liters correctly classified is similar, being 63.41% for SVMs and 85.76% for SDAs.

Meters with 0.1 pulse/liter precision

Preliminary results have also been obtained for 10 pulses/liter water meters. In this case there are only 19 counters available. Again, the results obtained with SDAs are better than those obtained with SVM. The mean accuracy of SVM and SDAs is 84.78% and 91.19%, respectively. In terms of liters, they fall to 73.47% and 85.89%.

Conclusion

The developed methodology allows the creation of models of neuronal networks or support vector machines for installations other than those analyzed in this work, provided there is available a period of training of previously classified events (manually, by an operator).

Also general models, based on the events recorded and classified in the Region of Madrid have been created, though their precision logically is poorer, and presumably it may drop considerably if these general models are transferred to another context of residential water use very different from where they have been trained.

Models specifically developed and trained using a large enough data set correctly classified, can achieve very good results in terms of precision. Methodology here depicted is actually in use for processing data from the *Panel for residential consumption patterns assessment and end-uses monitoring* project of Canal de Isabel II in Madrid.

References

[1] C. M. Bishop, Microsoft Research Ltd, "*Pattern Recognition and Machine Learning*", Springer Information Science and Statistics, 2006.

[2] Gleick, P. et al. (2003). *Waste Not, Want Not: The Potential for Urban Water Conservation in California*. Pacific Institute for Studies in Development, Environment, and Security

[3] W. DeOreo and P. Mayer, "*The End Uses of Hot Water in Single Family Homes from Flow Trace Analysis*", Aquacraft Inc. Boulder, Colorado, USA, 2001.

[4] Cubillo, F, Rueda, T., Ortega, S.: "*Microcomponentes y factores explicativos del consumo doméstico de agua en la Comunidad de Madrid*". Cuadernos I+D+I. Canal de Isabel II, 2008.

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

[6] 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.

[7] Scholkopf, B., & Smola, A. J. (2001). Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press.

[8] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.

[9] MIT Technology Review, Ten Breakthrough technologies 2013. http://www.technologyreview.com/lists/breakthrough-technologies/2013/.

[10] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P. A. (2010). "*Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion*". Journal of Machine Learning Research, 11 (Dec), 3371-3408

[11] Diaz Morales, R., & Vázquez, Á. N. (2016). Improving the efficiency of IRWLS SVMs using parallel Cholesky factorization. Pattern Recognition Letters, 84, 91-98.