Intelligent autonomous system for residential water end use classification: Autoflow

Author

Khoi, Anh Nguyen, Stewart, Rodney A, Zhang, Hong, Jones, Christopher

Published

Journal Title

Version

DOI

Copyright Statement

Downloaded from

Griffith Research Online

https://research-repository.griffith.edu.au

INTELLIGENT AUTONOMOUS SYSTEM FOR RESIDENTIAL WATER END USE CLASSIFICATION: *AUTOFLOW*

Authors:
Dr. Khoi Anh Nguyen
Griffith School of Engineering
Griffith University, Gold Coast Campus, Australia
Email: k.nguyen@griffith.edu.au

Associate Professor Rodney A. Stewart
Griffith School of Engineering
Griffith University, Gold Coast Campus, Australia
Queensland Australia 4222
Email: r.stewart@griffith.edu.au
Phone: +61 7 5552 8778

Associate Professor Hong Zhang
Griffith School of Engineering
Griffith University, Gold Coast Campus, Australia
Queensland Australia 4222
Email: hong.zhang@griffith.edu.au

Dr. Christopher Jones
City West Water
Sunshine, Victoria 3020, Australia
Email: Chris.Jones@citywestwater.com.au

Corresponding Author:
Associate Professor Rodney A. Stewart

Research Highlights

- Smart metering technology enables the capture of high resolution water consumption data
- Intelligent algorithms autonomously categorise single and combined water end use events
- Hybrid combination of HMM,ANN and DTW for pattern recognition problem
- Expert system developed to autonomously disaggregate water use into end use categories

Graphical Abstract
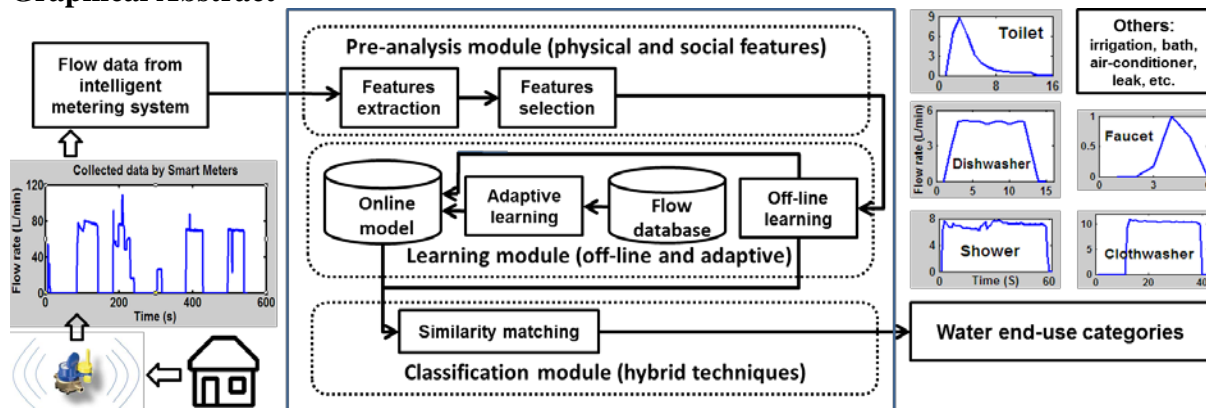
INTELLIGENT AUTONOMOUS SYSTEM FOR RESIDENTIAL WATER END USE CLASSIFICATION: *AUTOFLOW*

Abstract

Over half of the world's population will live in urban areas in the next decade, which will impose significant pressure on water security. The advanced management of water resources and their consumption is pivotal to maintaining a sustainable water future. To contribute to this goal, the aim of this study was to develop an autonomous and intelligent system for residential water end-use classification that could interface with customers and water business managers via a user-friendly web-based application. Water flow data collected directly from smart water meters connected to dwellings includes both single (e.g., a shower event occurring alone) and combined (i.e., an event that comprises several overlapping single events) water end use events. The authors recently developed an intelligent application called *Autoflow* which served as a prototype tool to solve the complex problem of autonomously categorising residential water consumption data into a registry of single and combined events. However, this first prototype application achieved overall recognition accuracy of 85%, which is not sufficient for a commercial application. To improve this accuracy level, a larger dataset consisting of over 82 thousand events from over 500 homes in Melbourne and South-east Queensland, Australia, were employed to derive a new single event recognition method employing a hybrid combination of Hidden Markov Model (HMM), Artificial Neural Networks (ANN) and the Dynamic Time Warping (DTW) algorithm. The classified single event registry was then used as the foundations of a sophisticated hybrid ANN-HMM combined event disaggregation module, which was able to strip apart concurrently occurring end use events. The new hybrid model's recognition accuracy ranged from 85.9 to 96.1% for single events and 81.8 to 91.5% for combined event disaggregation, which was a 4.9% and 8.0% improvement, respectively, when compared to the first prototype model. The developed *Autoflow* tool has far-reaching implications for enhanced urban water demand planning and management, sustained customer behaviour change through more granular water conservation awareness, and better customer satisfaction with water utility providers.

Key words: water end use event; residential water flow trace disaggregation; water micro-component; Hidden Markov Model; dynamic time warping algorithm; artificial neural network; water demand management

1. Introduction

Growing urban areas and climate change implications are raising concerns on water security (Sahin et al. 2014a). In response to growing pressure on urban water supplies, many governments and public utilities are investing significant funds in the development and implementation of strategies to ensure future urban water demand can be met (Sahin et al. 2014b). With the advent of advanced water metering, logging and wireless communication technologies, one of the top priorities in major metropolitan areas globally is to develop a smart water management system that helps to heighten water security through capturing, analysing, and disseminating near real-time water consumption data and concise reports to both water utilities and customers. Specifically, once in place, the system enables customers and utilities to actively monitor, through web-portal interfaces, real-time information about what, when, where and how water was consumed at their meter connection (e.g. 56 litre shower occurring between 06:55-07:08 on Tuesday 25 May 2014). It also allows individual consumers to log into their user-defined water consumption web page to view their daily, weekly, and monthly consumption summaries, as well as more granular charts displaying their water end-use patterns across major end use categories (e.g. leaks, clothes washer, dishwasher, tap, toilet, shower and irrigation). The analytical report generated by the new advanced integrated water management system will help utilities identify the water consumption patterns of their various consumer types and assist with a range of urban water planning and management functions (Stewart et al. 2010).

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 and can artificially influence water use behaviour (Froehlich et al. 2009, Nguyen et al. 2013b). 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 (i.e. replacement of traditional meter only). This approach uses pattern recognition (i.e. Hidden Markov Model and Dynamic Time Warping algorithm) coupled with other data mining techniques (i.e. event probability analysis) to automate the end use analysis process. A prototype software tool (*Autoflow*) was developed to provide a user-friendly platform to aid this process as described in the following paragraph.

The first prototype version of *Autoflow* was founded on an analytical procedure developed by the authors (see Nguyen et al. 2013a, b; 2014) and achieved an average overall pattern recognition accuracy across all categories of 85%. However, higher recognition accuracies and the ability to recognise and adapt to new or altered end uses are needed before a commercially viable software tool can be developed. 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 (i.e. currently below 70%).

This paper enhances previous research conducted by the researchers (Nguyen et al. 2013a, b; 2014) by combining Hidden Markov Model (HMM), Artificial Neural Network (ANN) and Dynamic Time Warping (DTW) techniques to examine both the shape pattern and physical characteristics of each event in order to identify the most likely end use category a particular flow pattern signature belongs. In this paper we have supplemented HMM with an adjunct ANN pattern recognition model in order to help capture water user habits in all human-dependent end uses and also the distinct operational features of all mechanised categories in that particular home. This has been achieved using ANN to decipher the likelihood occurrence for an extensive database of nine different characteristics or features of each end use event (i.e. volume, duration, maximum flow-rate, most-frequent flow rate, frequency of most frequent flow-rate, magnitude of initial flow-rate rise, magnitude of flow-rate drop at the end of event, gradient of initial flow-rate rise, and gradient of flow-rate drop at the end of the event). In this present paper, the influence of this newly supplemented ANN computational procedure has significantly improved the accuracy of the prior reported water end use event pattern recognition models for both the single (Nguyen et al. 2013a) and

concurrently occurring (Nguyen et al. 2013b) residential water end use events. Moreover, the compiled expert system software presented at the end of this paper is a demonstrably enhanced version of that presented in Nguyen et al. (2014) with significantly more accuracy and functionality. The present application is close to a commercial application standard.

The overall recognition process for the refined HMM-ANN-DTW method can be divided into four main stages: (1) likelihood estimation using HMM based on the event flow-rate pattern (i.e. flow pattern signature); (2) examination of physical feature similarities (e.g. rate of change of water flow) using ANN; (3) integration of HMM and ANN likelihood decisions to categorise events; and (4) where uncertainty still exists, apply further decision support algorithms such as DTW and event probability tables. The following sections of the paper will outline the background to this research, the establishment of classification model architecture, the customisation of HMM and ANN for the study, the model validation and calibration, and finally a description on how this novel research is an integral part of the future intelligent metering and information management future.

2. Background

2.1. Conventional water metering process

Water consumption readings are usually recorded manually on a quarterly or half yearly basis. Under most situations, a whole year's worth of water consumption data is described by only two to four data points in the water businesses billing system. Conventional water meters count each kilolitre of water as it passes through the meter and do not have the ability to record when (i.e., the time of day) and where the consumption takes place (e.g., washing machine, leaks) (Stewart et al., 2011). These systems produce limited and delayed water consumption information. The current water metering system does not typically provide real-time or continuous/frequent water consumption data, and in cases where it does, it does not provide a sufficient level of data resolution to allow water end-use event categorisation.

2.2. Existing autonomous water end use classification models

Along with the recent advancements of smart metering technology, several models have also been developed to automatically or semi-automatically classify the collected real-time or near real-time water consumption data into different categories. Nguyen et al. (2013a) provides a recent comprehensive critique on the strengths and weaknesses of each of the existing models. This critique, recommended that the utilisation of smart water meters at the property

boundary supported by firmware which could autonomously disaggregate water flow data into discrete water end use categories through applying advanced data-driven machine learning techniques was the most feasible approach to realise a vision of end use data being delivered to customers and utilities.

2.3. Data collection for the study

The database utilised in this study came from two recently completed end-use studies conducted in Brisbane (Beal and Stewart, 2011; Beal and Stewart, 2014) and Melbourne (Gan and Redhead, 2013), Australia, where high resolution water meters recording 0.014 L/pulse at five second intervals was available for over 500 homes in three years from 2010 to 2012. This data was manually disaggregated into end uses by analysts using the support of the Trace Wizard^{TM} software package. The outcomes of this end use analysis process are seven different sets of main water consumption patterns in residential homes, including shower, faucet (tap), dishwasher, clothes washer, toilet, bathtub, and irrigation. The number of samples used for model training and testing for each of these eight end use categories is detailed in Table 1 presented later.

2.4. Overview of utilised pattern recognition techniques

Intelligent metering technology coupled with advanced pattern recognition techniques enables a paradigm shift in the current level of water information provision available to the customer and water business. The aim of this current research project was to further enhance the recognition accuracy of the existing first prototype *Autoflow* model through applying various mathematical techniques. Among the more popular pattern matching and optimisation tools trialled included Refactoring method (Chang & Huang, 2012), Tabu search (Latif, 2013) and Optimisation using data kernel dependent technique (Zhang, 2014). Our exploratory trialling of these and other techniques revealed that a hybrid combination of HMM, ANN and DTW was the most suitable and accurate pattern recognition system.

HMM is a stochastic finite state automation defined by the parameter $\lambda = (\pi, a, b)$, where π is an initial state probability, a is state transition probability and b is observation probability, defined by a finite multivariate Gaussian mixture. Given an observed sequence $\mathbf{O} = (o_1, o_2, \ldots o_t, \ldots, o_T)$, a HMM model can be used to compute the probability of $\mathbf{O}$, denoted as $P(O|\lambda)$ and to find the corresponding state sequence (Q) that maximises the probability of $\mathbf{O}$,

denoted as P(Q|O, λ). The overall process for the establishment of a HMM model is presented in Section 3. Figure 1 illustrates an example of a left-right HMM model for an observation vector $\mathbf{O} = (o_1, o_2, o_3, o_4)$ which was employed in this study.
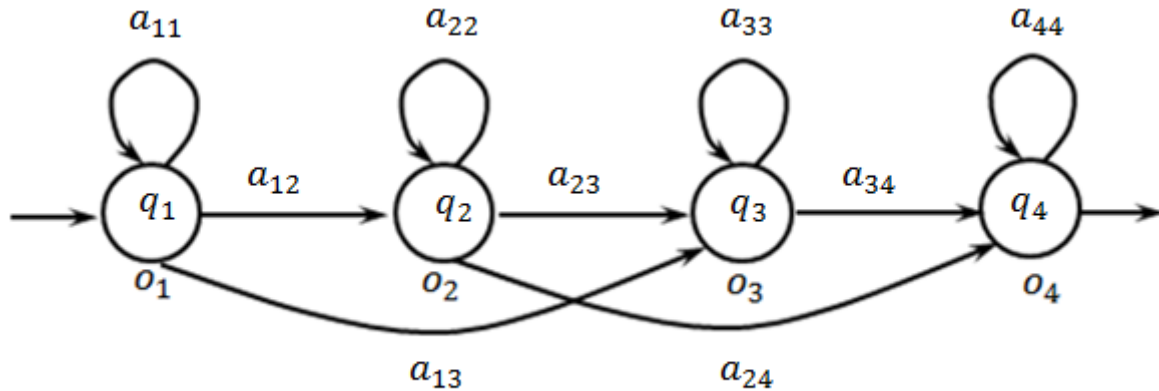
Figure 1 Example of left-right HMM model

HMM 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 (1993) and Tapia (2004). In this study, HMM was utilised as one of the classifiers for water end use classification decision making based on the event shape pattern. However, the weakness of this technique is that HMM does not adequately classify end use categories that are highly dependent on user behaviour; such end uses are highly variable meaning that they sometimes having features that closely resemble those in other categories. To illustrate this issue, Figure 2 shows that three end use event categories (i.e. shower, bathtub and irrigation) can 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 differentiate between them.
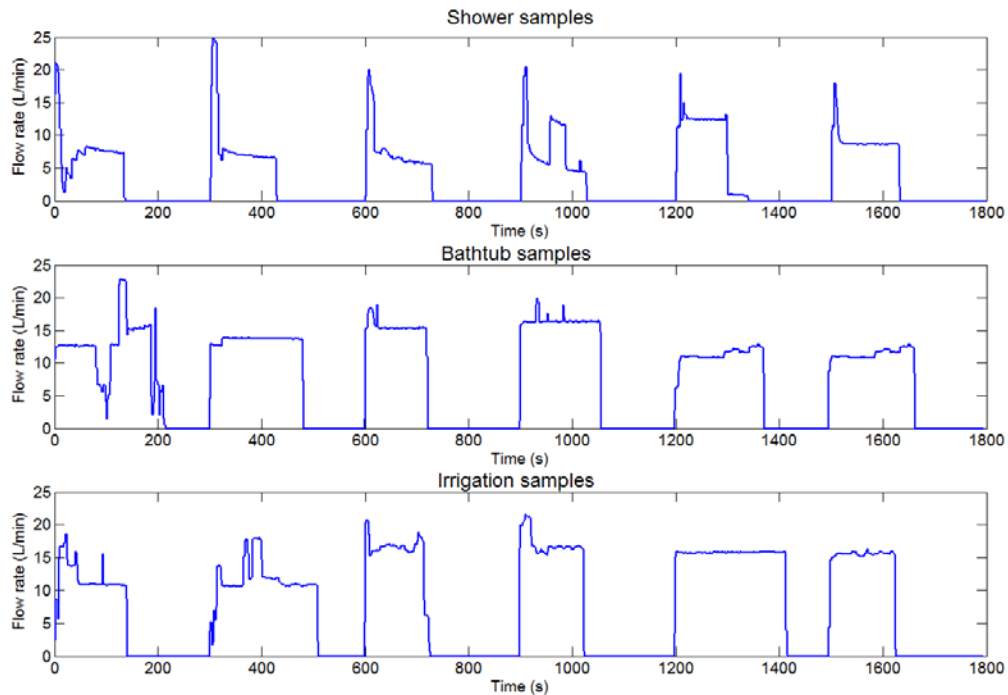
Figure 2 Similar samples extracted from three different end-use categories

To overcome the above issue, another technique called Artificial Neural Network (ANN) was employed. 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 in terms of physical characteristics (e.g. volume, duration, maximum flow rate, etc.).

The last applied mathematical tool is the Dynamic time warping (DTW) algorithm, which is a popular method for measuring the similarity between two time series of different lengths. In general, this task is performed by finding an optimal alignment between two series with certain restrictions. The sequences are extended or shortened in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension (Myers and Rabiner, 1981). The goal of DTW is to find a mapping path which has the minimal mapping distance.

This technique has been widely applied in prototype selection (e.g. Nguyen et al., 2011), pattern recognition (e.g. Myers and Rabiner, 1981; Muller, 2007; Rabiner and Juang, 1993; Sakoe and Chiba, 1978; Manmatha and Srimal, 1999; and Marquez, 2001) or word image searching (Manmatha and Rath, 2002). DTW played an important role in this study as it was utilised for the task of searching for linked cycles of water use related to one particular end use event for mechanised end use events (e.g. clothes washer and dishwasher) that were misclassified by HMM and ANN. In essence, clothes washers and dishwashers have patterns of cycles of water use associated with a particular customer 'wash' selection, which can be recognised using DTW.

In summary, the refined model formulated through this current study has overcome the limitations of the first version of *Autoflow* reported in Nguyen et al. (2014) by using a hybrid ANN-HMM approach to better handle the variability of user-dependent patterns such as shower, bathtub and irrigation. Specifically, this problem was successfully resolved by combining HMM to help inspect the flow rate pattern and ANN to investigate the event physical characteristics. The integration of analysis outputs from these two techniques has meant that more robust decisions could be made for the more complex water end use categories.

3. Classification model architecture

3.1. Overview

With the available database, the disaggregation process of water end use events from the raw data was developed. As mentioned previously, single events are defined as those which occur in isolation (e.g. toilet flushing only), while combined events have simultaneous occurrences of water usage (e.g. a shower occurring while someone else is using a tap). Combined events are more challenging to disaggregate into discrete single events. As the very first step of the

classification process (Figure 3), the HMM algorithm is used to recognise if an event can be clearly allocated to a particular single event category or is most likely a combined event. Remaining events from this process are placed into two groups, namely, unclassified single events and combined events. In the case of unclassified single events, a combination of HMM, ANN and DTW was then employed to assign them to appropriate end use categories. The detailed procedure is shown in Figure 4. The next important task involves the combined event classification, which is one of the most complicated problems in the field of pattern matching. The detailed procedure is summarised in Figure 8.
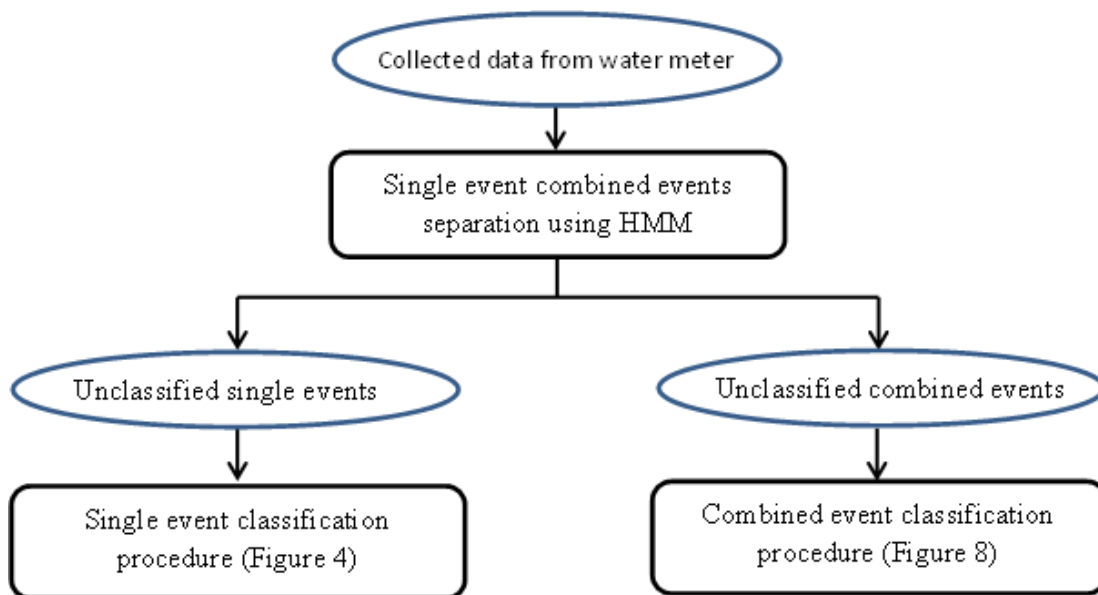
Figure 3 Overview of the classification process

3.2. Single event analysis

The single event analysis module using HMM in the *Autoflow* prototype software version 1 has proven to be very suitable for classifying mechanised water end use categories such as clothes washer, dishwasher and toilet (Nguyen et al. 2013a and 2013b). However, lower levels of accuracy were accomplished when adapting this technique on complicated shower, bathtub and irrigation events as their flow rate patterns and corresponding physical characteristics vary considerably (Nguyen et al. 2013a and 2013b). This study has found that 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. 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.
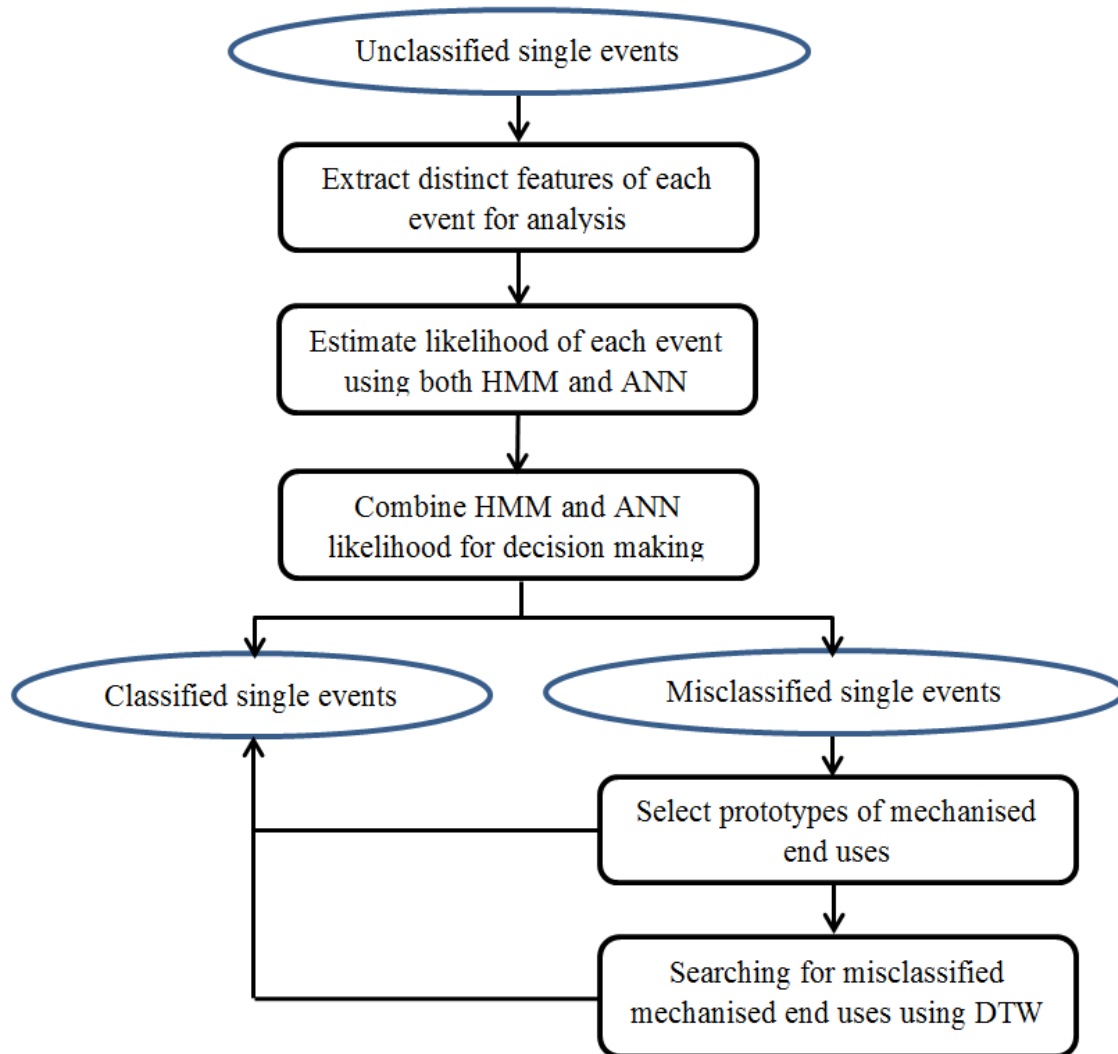
Figure 4 Single event classification procedures

As shown in Figure 4, to enable a classification process using ANN, distinct physical features, explained in Section 4.2, from each event have to be initially extracted and used as the main input for the classification process. The likelihood estimation of each event using HMM and ANN is then conducted, which allows the final decision to be made.

At the end of this step, most unclassified single events will be assigned to appropriate categories. However, even after these procedures are completed, there are often a small number of misclassified events due to their complicated patterns or highly similar patterns to other end use categories making it difficult to assign them. To capture the majority of these unclassified events a further analysis procedure was developed, which essentially searched for wash-cycle patterns evident in the mechanised end use categories (e.g. clothes washer).

11

The searching process starts with selecting clothes washer and dishwasher prototypes from the already classified events of these two categories. Figure 5 below presents a group of classified clothes washer events from the testing of an independent new data in Melbourne, from which the new prototypes (Figure 6) were then extracted by using the prototype selection algorithm developed by (Nguyen et al., 2011). With the availability of the representative clothes washer samples, the searching can be performed by looking for any event that has similar shape to these prototypes using DTW. An overall model verification has shown that this additional searching process has helped increase the classification accuracy of these mechanised end uses by 1-3%.
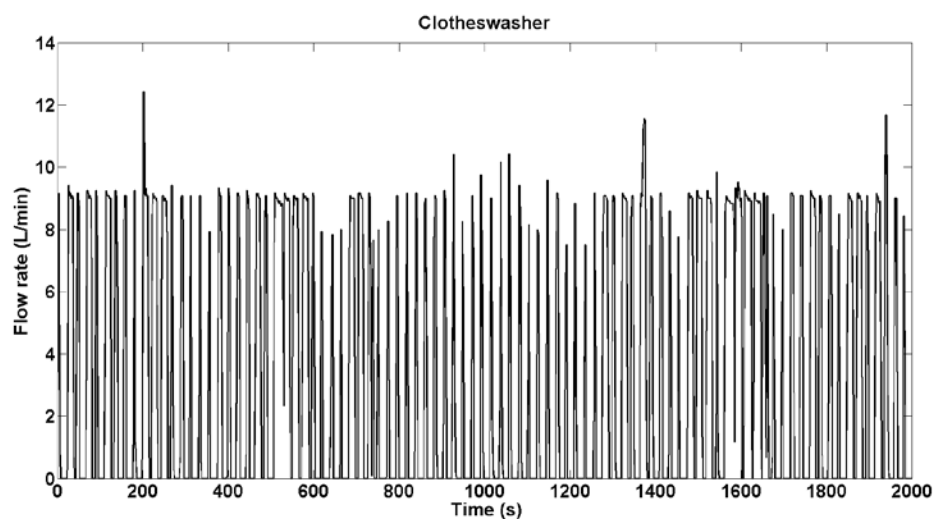
Figure 5 Classified clothes washer events
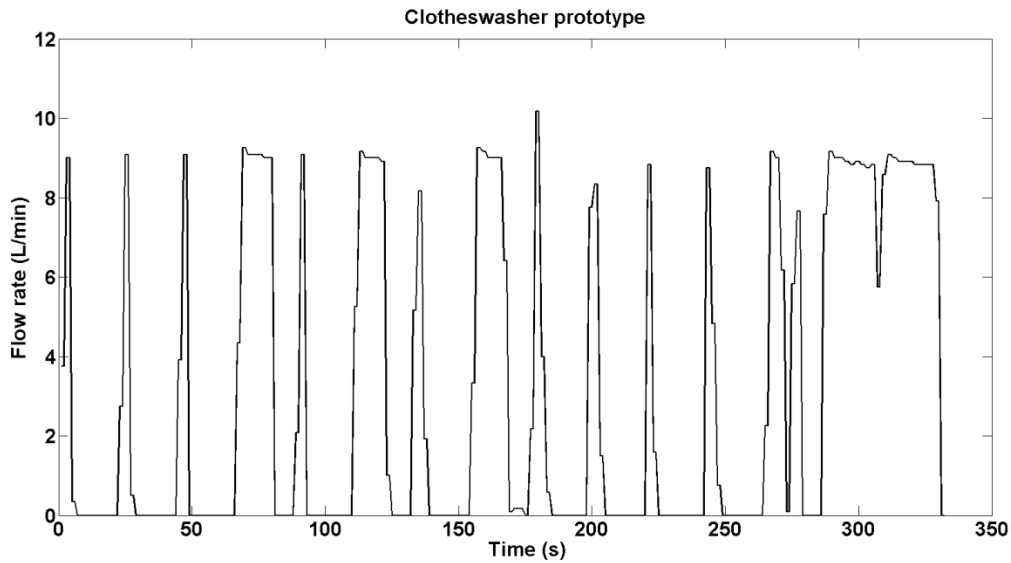
Figure 6 Prototypes of clothes washer events in the tested property

3.3. Combined event analysis

The overall procedure for combined event disaggregation is detailed in Nguyen et al. (2013b). Therefore, only a brief summary of the techniques applied for this module is presented herein. Some enhancements have been made to the existing combined event module to improve classification accuracy; these have been elaborated on in greater detail.

A combined event is formed by at least two simultaneously occurring single events. There is no restriction on the starting and finishing time of each discrete event making up a combined event, as long as they have an overlapped period with each other. An intensive analysis of the collected data reveals that most combined events last more than one minute, allowing enough time for many end-use categories to overlap. In a combined event, the longest component is named the "base-event", and all other shorter events are called "sub-events", which are superimposed on top of the base event (i.e. in Figure 7, Event 1 is defined as a base event, and all other events are named sub-events). An important finding is that most of the collected base events include the shower, bathtub and long irrigation, as they all last substantially longer than the other, shorter end-uses. For sub-events, the most common categories are the tap, clothes washer, dishwasher, toilet, and short irrigation.

The established HMM-ANN-DTW hybrid model for single event recognition also plays a major role for combined event disaggregation along with other pre-determined criteria. The

whole combined event analysis process is separated into two main stages: (1) Sub-event analysis; and (2) Base-event analysis (Figure 8).

In the first stage of analysis, a separation process employing the modified gradient vector filtering method is applied to disaggregate the uncategorised combined event into one base sample and several sub samples, where the term "sample" is used to refer to the products obtained from the separation process before the classification. Once these samples are assigned to their proper categories, they are called "events". This analysis section includes two layers. The HMM method is applied to the sub samples to determine whether they are actual complete single events or just parts of other events within the combined event.

At this stage, to achieve a reliable decision related to the sub-event judgment, some pre-determined criteria are incorporated into the existing HMM model. The criteria were established from an intensive statistical study on the entire database of end use events, which reflects the true user's behaviour on different types of water end-use categories. The products obtained after the first layer analysis are the actual, already- classified, single events, and sometimes some additional unclassified sub samples, which do not satisfy the threshold criteria to be assigned to any particular water end-use category. These undetermined samples are then passed through the second layer of analysis which is similar to the first one. The only difference between Layer 1 and 2 is that all products achieved after the separation process in Layer 2 (both sub and base samples) would be classified using criteria applied only for sub-events. The unclassified samples after sub-event Layer 2 analysis, where they exist, are considered as parts of the original base event and, therefore, are returned to that base event at the same time period they existed at prior to the commencement of the separation process.
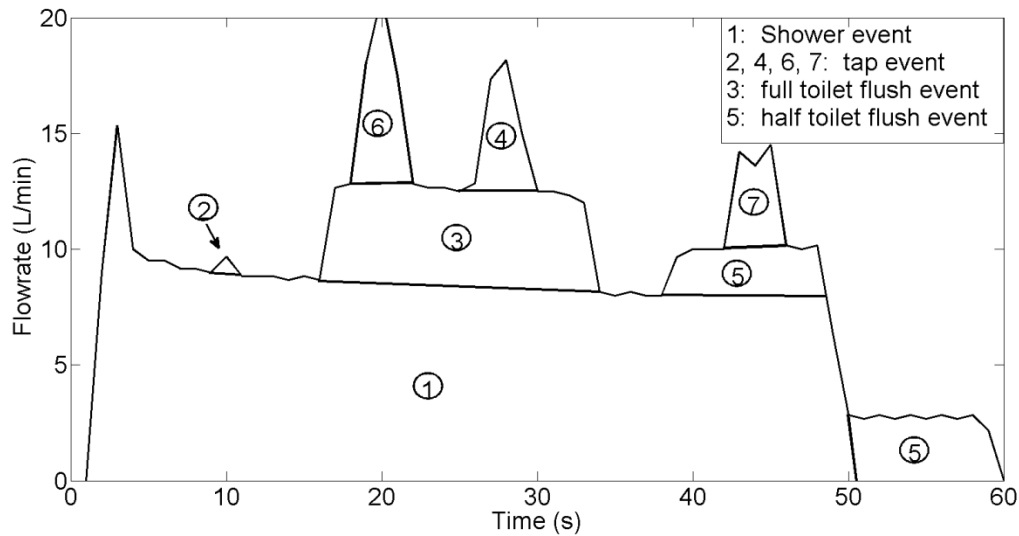
Figure 7 Example of a typical combined event (Nguyen et al, 2013b)

A base event, as previously defined, is the longest single event within the combined event. However, in the second stage of the analysis, the classification of the base sample, achieved after the initial separation process, often remains problematic; that is, the decision needs to be made whether this event is now a single event or another combined event. This uncertainty arises because it is just the remaining product after small or spiky sections are taken away from the original combined event. To tackle this issue, the subjected base sample is dissected into many smaller parts, using the same gradient vector filtering technique, for further analysis. The outcome of the second stage analysis, following the HMM classification process, is a classified a single base event, with the potential for other classified sub events, where they exist.
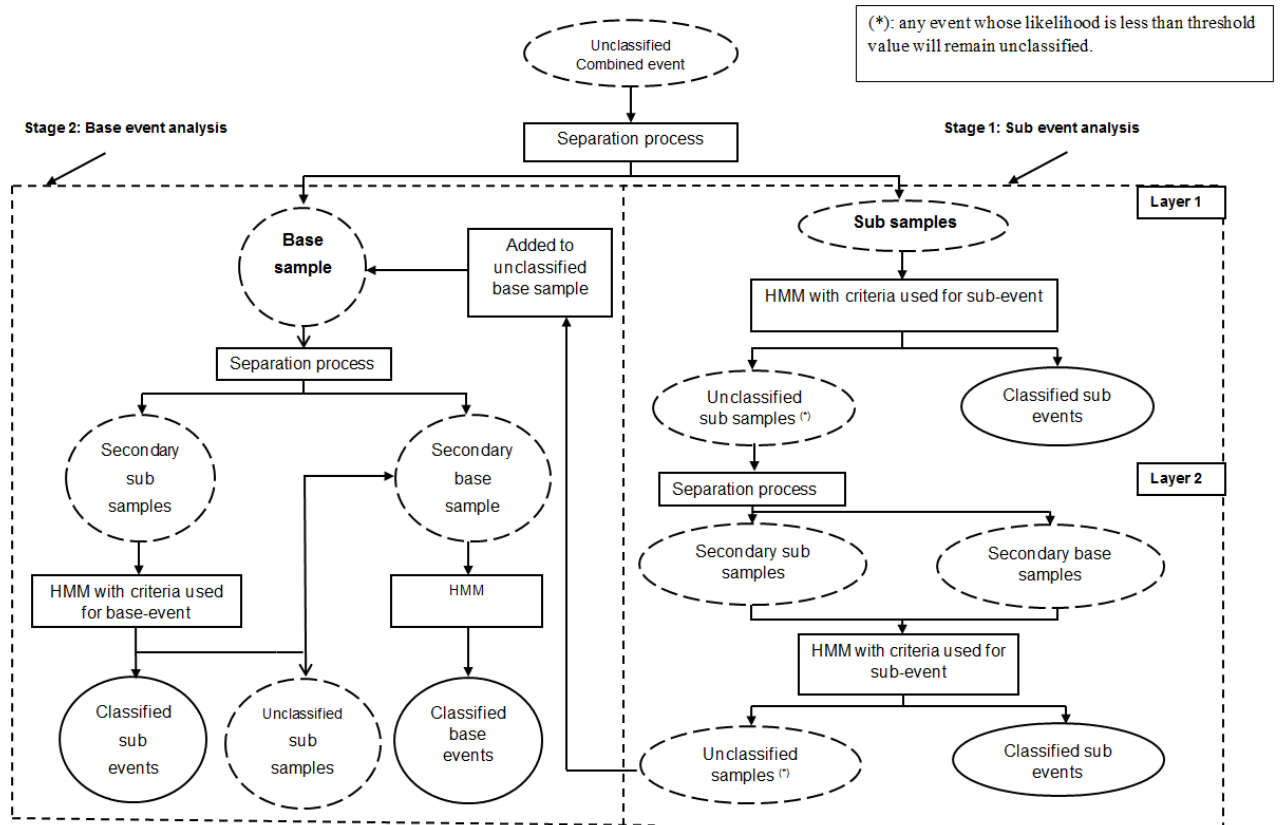
Figure 8 Combined event analysis process

In the above combined event analysis procedure, the critical task relates to the classification process using HMM as it directly assigns unknown samples to appropriate categories. Therefore, significant improvement would be achieved if this analysis module is modified to better reflect the event patterns of the currently tested property. In the present study, the combined event module was enhanced by the addition of a new analysis procedure which helped to significantly improve the recognition accuracy. This new procedure involved the creation of automatically customised HMM and ANN models for the tested home from the classified single events in that home in order to help capture water user habits in all human-dependent end uses and also the distinct operational features of all mechanised categories in that particular home.

4. Enhanced water end use classification techniques

The conceptualisation of the architecture for the model described above was crucial before the operationalization of the pattern matching algorithms and associated analysis processes. This section briefly describes the model operationalization process including a brief

introduction of the selected HMM, ANN and DTW techniques, as well as their customisation to the problem of water end use event classification..

4.1. HMM customisation

As presented in (Nguyen et al. 2013a, b and 2014), HMM was applied as the main technique to learn the shape pattern of all collected events. 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 Equations 1 to 8:

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 Nguyen et al (2013a) 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)

$$\alpha_t(i) = P(o_1 o_2 ... o_t, q_t = i \mid \lambda) \tag{1}$$

- $\beta_t(i)$: 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 ($q_t = i$)

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

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

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)} \tag{3}$$

- $\xi_t(i,j)$: 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 .

$$\xi_t(i, j) = \frac{P(q_t = i, q_{t+1} = j, \mathbf{O} \mid \lambda)}{P(\mathbf{O} \mid \lambda)} \tag{4}$$

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

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

Step 3: Calculate the following parameters for each water flow sequence ($\mathbf{O}$)

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

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

($\mathbf{O}$)

18

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

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 Equations 5 – 7:

$$\bar{\pi}_i = \gamma_1(i) \tag{5}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{6}$$

$$\bar{b}_j(o_k) = \frac{\sum_{\substack{t=1 \\ \text{such that } \mathbf{o}_t = \mathbf{v}_k}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \tag{7}$$

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 ($\mathbf{O}$) 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:

$$P(\mathbf{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{8}$$

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).

At the end of the HMM training process, eight HMM models were established with the first one determining whether the tested event occurred in isolation or was a group of concurrent events, and the remaining seven models being for discrete single events within the major residential water end use categories, including shower, faucet, clothes washer, dishwasher, toilet, bathtub and irrigation.

4.2. ANN model customisation

On the other hand, to facilitate the end use classification using ANN, 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 (e.g. Trace Wizard), the inclusion of the last four can also be considered as critical features that can improve recognition accuracy. These features are defined below and are also illustrated in Figure 9:

(vi) *Magnitude of initial flow-rate rise (Δf_1)* is defined as the flow-rate rise at the initial phase of the event when the consumer starts using water.

(vii) *Magnitude of flow-rate drop at the end of event (Δf_2)* is the flow-rate drop at the end phase of the event when the valve is shut.

(viii) *Gradient of initial flow-rate rise (g_1)* is determined by getting Δf_1 divided by the time it takes to reach this flow-rate (Δt_1).

$$g1 = \Delta f_1 / \Delta t_1 \tag{9}$$

(ix) *Gradient of flow-rate drop at the end of event ($g2$)* is determined by getting Δf_2 divided by the time it takes for the flow to revert to zero (Δt_2).

$$g1 = \Delta f_2 / \Delta t_2 \tag{10}$$
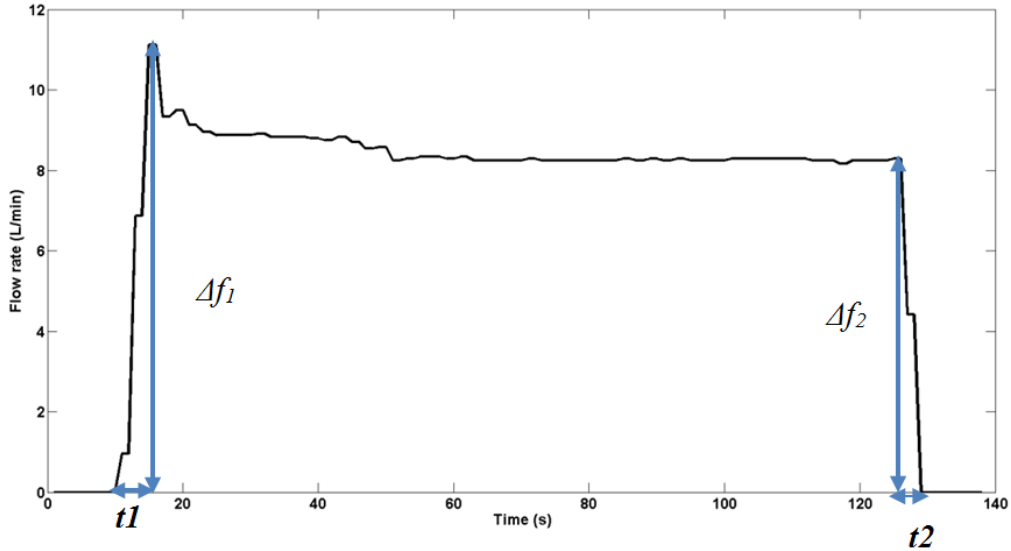
Figure 9 Illustration of features (vi) to (ix) used in ANN training process

Once the required features of all events have been obtained, the ANN training process using the back-propagation algorithm can be started following the network presented in Figure 10. It should be noted that as the final classification decision will be made based on the likelihood obtained from both HMM and ANN models, a training algorithm using gradient descent method that requires least computational cost while still being able to capture major pattern features has been selected for this study. The output of this process is an ANN model that is able to turn any unclassified event into one of the seven end use categories, namely shower, faucet, clothes washer, dishwasher, toilet, bathtub, and irrigation.
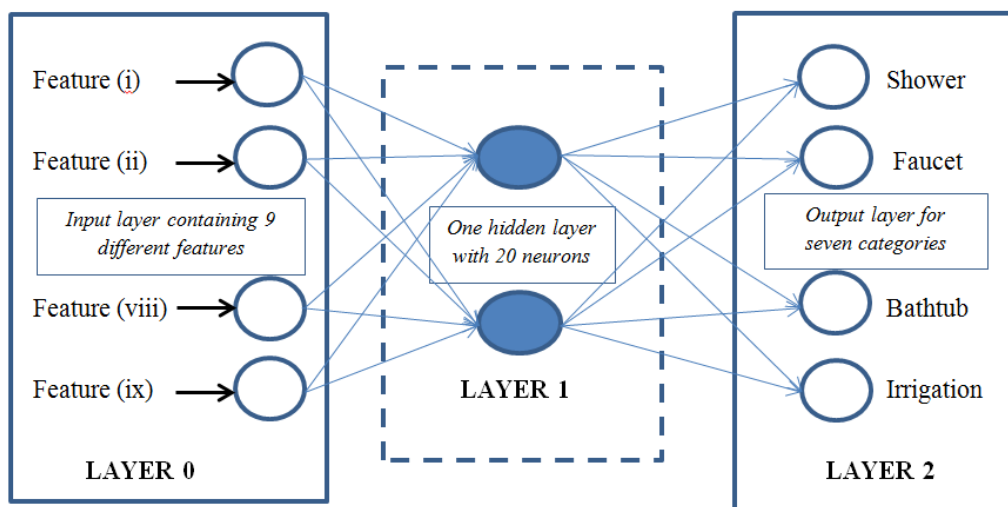
Figure 10 Model for ANN training

According to Figure 10 that illustrates the multilayer network with 2 layers, N_n represents the number of neurons in the nth layer. Here, the network is presented the pth pattern of training sample set with 9-dimensional input $X_p = (X_{p1}, X_{p2}, \ldots, X_{p9})$, and 7-dimensional known output response $T_p = \{T_{p1}, \ldots, T_{pi}, \ldots, T_{p7}\}$. The actual response to the input pattern by the network is represented as $O_p = (O_{p1}, \ldots, O_{pi}, \ldots, O_{p7})$. Let Y_{jn} be the output from the jth neuron in layer n for p^{th} pattern; W_{njk} be the connection weight from k^{th} neuron in layer $(n-1)$ to j^{th} neuron in layer; and δ_{nj} be the error value associated with the jth neuron in layer n. The training process for the proposed ANN model can be summarised as follows:

Step 1: Initialise connection weights into small random values between 0 and 1.

Step 2: Present the pth input vector of category i which is in the form of $X_p = (X_{p1}, X_{p2}, \ldots, X_{p9})$, and the corresponding target vector $T_p = (0, \ldots, T_{pi} = 1, \ldots, 0)$. It should be noted that in vector T_p, the value at i^{th} position is 1 and other values are 0 where $1 \leq i \leq 7$ as the total category number in the study is 7.

Step 3: Pass the input values to the hidden layer, which contains 20 neurons. For every input neuron j in the input layer (named as layer 0), perform:

$$Y_{0j} = X_{pj} \tag{11}$$

Step 4: For every neuron j in layer $n = 1$ and $n = 2$, find the output from the neuron:

$$Y_{nj} = f\left(\sum_{k=1}^{Nn-1} Y_{(n-1)k} W_{njk}\right), \tag{12}$$

where

$$f_x = \frac{1}{1+\exp(-x)} \tag{13}$$

Step 5: Obtain output values. For very neuron j in the hidden layer, perform:

$$O_{pj} = Y_{2j} \tag{14}$$

Step 6: Calculate error value δ_{1j} for every neuron j in the hidden layer, followed by weight adjustment. For the output layer, the error value is:

$$\delta_{2j} = Y_{2j}(1 - Y_{2j})(T_{pj} - Y_{2j}) \tag{15}$$

and for the hidden layer:

$$\delta_{1j} = Y_{1j}(1 - Y_{1j})\sum_{k=1}^{7} \delta_{2k} W_{2kj} \tag{16}$$

The weight adjustment can be completed from neuron k in layer $(n - 1)$ to every neuron j in layer n.

$$W_{njk,new} = W_{njk} + \beta \delta_{nj} Y_{nj} \tag{17}$$

where β represent the normalised adjustment factor. In this study, after considering the computational time and achieved accuracy, β was selected as 0.1 and the overall training procedure will be terminated when the weight adjustment (i.e. gradient) drops below 10^{-5}. It should be noted that to achieve the error value as presented in Equation 15, a sum-of square error function was applied (Equation 18):

$$E = \frac{1}{2}\sum_{p=1}^{9}\sum_{i}^{7}(T_{pi} - O_{pi})^2 \tag{18}$$

With the availability of the customised HMM and ANN model, the likelihood estimation of each event using these two techniques is conducted following the below logic:

Given $A = (a_1, a_2, \dots, a_7)$ 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, $B = (b_1, b_2, \dots, b_7)$ being the achieved likelihood when estimating this event with ANN, and $C = (c_1, c_2 \dots c_i, \dots c_7)$ 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 C.

## 4.3. DTW model for mechanised event searching

As mentioned prior, DTW was applied as a supporting technique to search for events from mechanised end used categories (e.g. clothes washer and dishwasher) that were misclassified by the hybrid ANN and HMM model in the first round of the analysis. This task could be achieved by using DTW to firstly select a certain number of prototypes from the classified

clothes washer and dishwasher events, then comparing those prototypes against the remaining collected events to find the similar patterns. The procedure and constraints of the DTW algorithm are summarised as:

(i) Given two vectors $\mathbf{p} = (p_1, \ p_2, \ldots p_i, \ldots p_m)$ and $\mathbf{q} = (q_1, q_2, \ldots q_j, \ldots, q_n)$ of length m and n respectively as in Figure 11. The boundary condition is that the two end pairs of the vectors are matched together, i.e. (p_1, q_1), (p_m, q_n).

(ii) Define $d(i, j) = |p_i - q_j|$, with (i, j) is a given node in the mapping path, the possible fan-in nodes are restricted to $(i-1, j), (i, j-1), (i-1, j-1)$. This local constraint guarantees that the mapping path is monotonically non-decreasing in its first and second arguments. Moreover, for any given element, it should be able to find at least one corresponding element, and vice versa.

(iii) Define $D(i, j)$ as the accumulated DTW distance between points of $\mathbf{p}$ (p_1 to p_i) and points of $\mathbf{q}$ (q_1 to q_j), where $(1 \le i \le m)$ and $(1 \le j \le n)$ and $D(i, j) = d(i, j) + \min\{D(i-1, j), D(i-1, j-1), D(i, j-1)\}$, with initial condition $D(1,1) = d(1,1)$.

(iv) Therefore, the final accumulated DTW distance between $\mathbf{p}$ and $\mathbf{q}$ is $D(m, n)$.
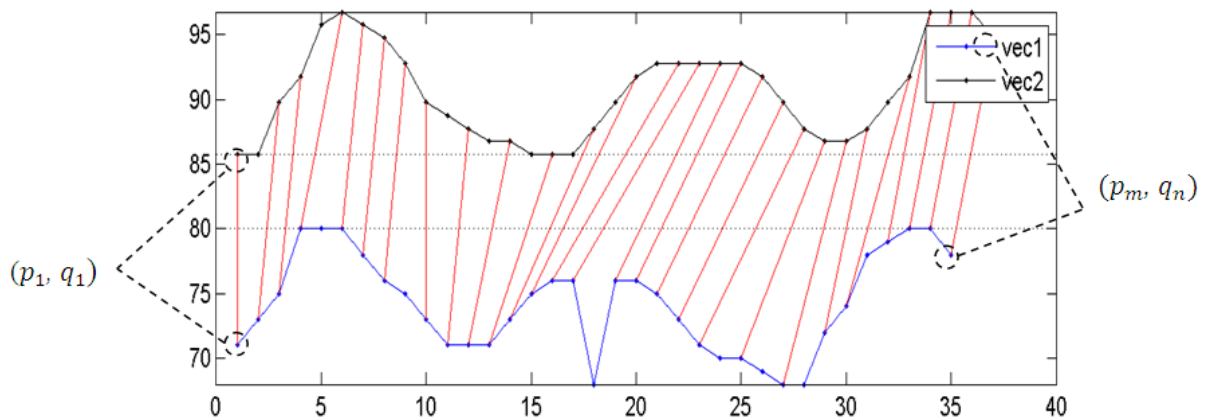
Figure 11 Boundary condition for the two end points in DTW path mapping (Nguyen et al, 2014)

5. Model calibration

As there is no particular rule for the selection of the number of hidden layers and neurons for each layer of an ANN model, and also given that the developed Autoflow system only needs

to classify a limited number of end use category (i.e. 7 in this study and is likely to be less than 12 in future version), the adoption of an ANN structure with 20 neurons and one hidden layer is considered as an ideal solution for this problem.

This selection has been achieved after several intensive verification processes when considering both analysis time and the obtained accuracy. In the future Autoflow version when new end use categories are introduced into the classification process, an overall validation of the existing ANN structure is recommended to determine whether the adoption of 20 neurons and 1 hidden layer is still effective; however, it is expected that no change would be made as according to one of the most reliable empirically-derived rules, one hidden layer is sufficient for the large majority of problems and the optimal size of the hidden layer is usually equal to the total size of the input and the output layers (Heaton, 2008).

In terms the developed HMM model, the selection of number of states was based on the pattern complexity and disregard the end use category the pattern belong to. In this study, 100 states was obtained to help model complicated events lasting up to 10 hours, which is very unlikely to occur in reality, therefore, no modification is required for the HMM model in all circumstances.

6.   Model verification

As described below, three stages of model testing and validation were undertaken in this study:

(i)   Model testing utilised HMM, ANN and the herein recommended hybrid combination of ANN-HMM on a total of 16,000 untrained single event samples.

(ii)   Single event model verification was conducted by applying the *Autoflow* version 1 (see Nguyen et al, 2013a) and 2 (i.e. herein presented enhanced model) software tool to three independent homes equipped with high resolution smart meters. These independent homes were not included in the training or testing datasets, and each had supporting diaries to confirm each and every end use event.

(iii)   The same validation process was conducted as is described in (ii) but for a combined event.

6.1. Single event analysis module testing on reserved data

16,000 samples (20% reserved data) were used to illustrate the accuracy achieved for the standalone HMM and ANN models as well as the enhanced hybrid HMM-ANN model proposed herein. End use event classification accuracy was calculated by dividing the number of correctly classified events for each category by the number of events for testing. For example, for the classification of shower events when using HMM alone (Table 1), the number of testing samples was 3,000 and the number of events correctly identified was 2,109, which resulted in an accuracy of 70.3%. Overall accuracy for all categories for each tested model (i.e. HMM, ANN, HMM_ANN) was calculated by summing the total number of accurately identified events (i.e. 12928, 12464 and 14864 events for the HMM. ANN and HMM+ANN models respectively) and divided by the total number of testing samples (i.e. 16,000). 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 recognition (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 consistent distinct patterns which can be correctly identified by HMM alone. Overall, the newly developed hybrid method has demonstrated that it is superior to the prior method which used HMM as the sole classifier when testing on untrained samples.

Table 1 Single event 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	3,000	70.3	78.6	93.8
2. Faucet	21,985	3,000	76.2	71.7	90.8
3. Clothes washer	15,211	3,000	82.6	72.8	91.7
4. Dishwasher	13,342	3,000	90.9	76.9	96.0
5. Toilet	15,222	3,000	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	80.8	77.9	92.9

6.2. Verification process for single event analysis module on independent homes

HMM was selected as the core classifier in *Autoflow* Version 1, which was assisted by the utilisation of DTW and different probability functions to refine the classification through a post-processing procedure (Nguyen et al. 2013a). The *Autoflow* Version 2 single event analysis module integrates ANN and other recognition enhancements. For model verification, three totally independent homes that were retrofitted with high resolution smart meters were recruited and their data logged over a 1-2 week period. These households were also incentivised to provide a very detailed water consumption diary to ensure that a near perfect registry of correctly classified water end use events could be created to ensure that *Autoflow* (version 2) model validation could be completed. Table 2 displays the recognition accuracy achieved for the current and predecessor *Autoflow* software tool for the three independent homes.

Table 2 Comparison of the single event analysis module used in *Autoflow* Version 1 and 2

End use category	Home 1		Home 2		Home 3	
	Version 1 (%)	Version 2 (%)	Version 1 (%)	Version 2 (%)	Version 1 (%)	Version 2 (%)
Tap	83.2	87.2	95.9	94.2	80.1	89.5
Dishwasher	92.1	93.5	86.5	88.0	96.5	96.5
Clothes washer	91.9	92.1	86.3	90.1	93.2	94.1
Shower	84.8	89.2	94.8	94.8	94.5	95.6
Toilet	78.9	83.5	80.0	86.3	90.3	89.5
Irrigation	NA	NA	56.4	80.6	66.1	80.2
Bathtub	71.9	85.6	NA	NA	70.7	79.9
All categories	84.1	**89.2**	84.1	**90.5**	85.7	**90.9**

As detailed in Table 2, *Autoflow* Version 2 was shown to be a superior classification system. For end use categories which had clearly defined patterns such as dishwasher and clothes washer, an average classification accuracy increase of 1.3% and 1.6%, respectively (average of three homes) ,was achieved. These relative small increases in event classification accuracy is due to the fact that the HMM technique that was used as the core classifier in Version 1 of

Autoflow is sufficiently robust for classifying consistent mechanised water flow patterns that are less influenced by human behaviours.

For end use categories that are more behaviourally influenced (i.e. tap, shower, toilet, bathtub and irrigation) and thus had higher flow pattern variability, there were much greater improvements in classification accuracy when the additional analysis processes were applied. An accuracy increase achieved considering the average of the three homes was 3.8% for tap, 1.8% for shower, 3.4% for toilet, 11.5% for bathtub and 19.2% for irrigation. This improvement in classification accuracy indicates that the enhanced analytical methods underpinning Version 2 of the *Autoflow* tool, which also utilised ANN as a key classifier, is more effective in recognising events than Version 1.

6.3. Verification process for combined event analysis module on independent homes

The new combined event analysis model utilised in Version 2 was verified on the same three independent homes described above. Table 3 shows the total number of events for each end use category that were components of the 25 combined events stripped apart by the formulated method, the number of correctly classified events using both Version 1 and 2 of *Autoflow*, and the corresponding accuracy of each version.

Table 3 Comparison of old and new methods in combined event analysis

Event category	Number of events for verification	Version 1	Accuracy (%)	Version 2	Accuracy (%)
Shower	12	10	83.3	10	83.3
Faucet	59	48	81.3	54	91.5
Clothes washer	13	10	76.9	11	84.6
Dishwasher	11	9	81.8	9	81.8
Toilet	14	11	78.5	12	85.7
Bathtub	7	5	71.4	6	85.7
Irrigation	6	4	66.7	5	83.3
Total	122	97	79.5	107	87.7

Table 3 shows that the disaggregation accuracy using the new model is over 80% for all water end-uses, with the highest being 91.5% for faucet, and the lowest being 83.3% for irrigation. The classification accuracy of complicated end uses such as bathtub and irrigation were also relatively high at 85.7% and 83.3%, respectively. This improvement in recognition accuracy was due to the utilisation of a hybrid combination of HMM and ANN models that were automatically customised for these particular homes using self-learning functions. In comparison with *Autoflow* Version 1, an accuracy improvement was achieved for most of the end uses, such as faucet (81.3% to 91.5%), clothes washer (76.9% to 84.6%), toilet (78.5% to 85.7%), bathtub (71.4% to 85.7%) and irrigation (66.7% to 83.3%). Although a relatively small number of combined samples were utilised for the verification process, the increasing trend in the obtained accuracy demonstrates that the refined model offers significantly better recognition performance.

7. Residential water end use categorisation software application – *Autoflow* Version 2

7.1. Application outputs

Through integrating and codifying the analytical processes contained in the single, combined and adaptive learning modules, the Autoflow Version 2 could be formulated that would be able to autonomously categorise remotely collected residential water consumption data received from smart meters into a repository of end use events with the achieved accuracy higher than that of the first version. This software application, developed using user-defined and built-in functions in Matlab programming language, offers different types of results presentations. Figure 12 shows the main interface, which provides important information to the customer, such as a summary of the classified volume of each end-use category during a specific period of time, which is supported by a detailed description of the start time, end time, volume, duration, maximum flow rate and most frequent flow rate of each classified event.
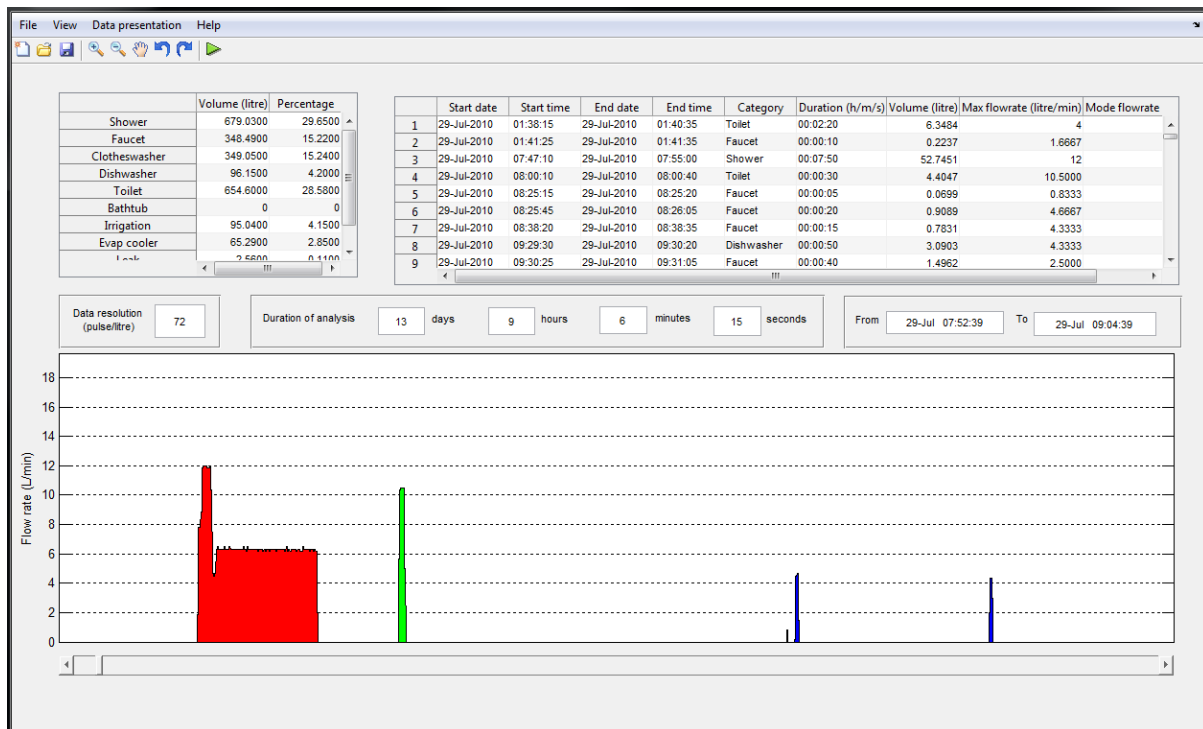
Figure 12 Main interface of *Autoflow*

Apart from the above information, a list of descriptive statistical data on each end use category is also provided, which include the average and mean values of volume, duration and flow rate extracted from the classified events. Another useful output of a water end use study is the daily end use diurnal demand graph. This graph can be automatically created from the repository of classified end use events, and is highly beneficial to both consumers and water businesses seeking to better understand how residential water consumption is being used, at an end use level, across various significant days of the year (i.e. average weekday, average weekend day, peak day, average day peak month). This data is particularly useful for water infrastructure planning (e.g. water pipe network augmentation planning) as it informs network modelling engineers of the peak demand flow rates as well as the key end uses contributing to that peak demand (i.e. evening shower use combined with clothes washer contributes to morning peak).

7.2. *Autoflow* computational efficiency

The developed *Autoflow* system was tested on a 2-week dataset (0.014 L/pulse every 5s data resolution) collected from160 homes located in South-east Queensland Australia in order to evaluate its computational efficiency. Using a standard desktop computer processor (Duo-core 2.4 GHz), the processing time to analyse these 160 residential customer connections and their associated individual water end use events was determined (Figure 13).
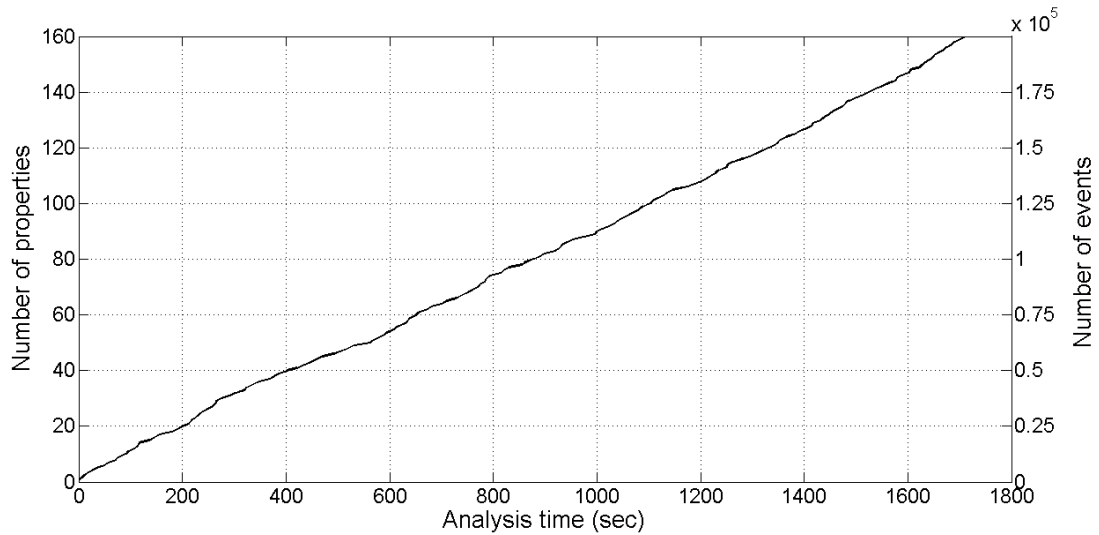
Figure 13 Computational efficiency of *Autoflow* application on Duo-core 2.4 GHz processor

Figure 13 illustrates that *Autoflow* has significantly reduced the flow data end-use disaggregation analysis time from 2 hours using the previously described conventional semi-manual method to approximately 10 seconds for a two-week dataset. In terms of number of events, *Autoflow* can pattern recognise almost 100 events every second, which is almost 500 times faster than an expert using *Trace Wizard*; the current most popular software on the market to achieve a similar level of accuracy.
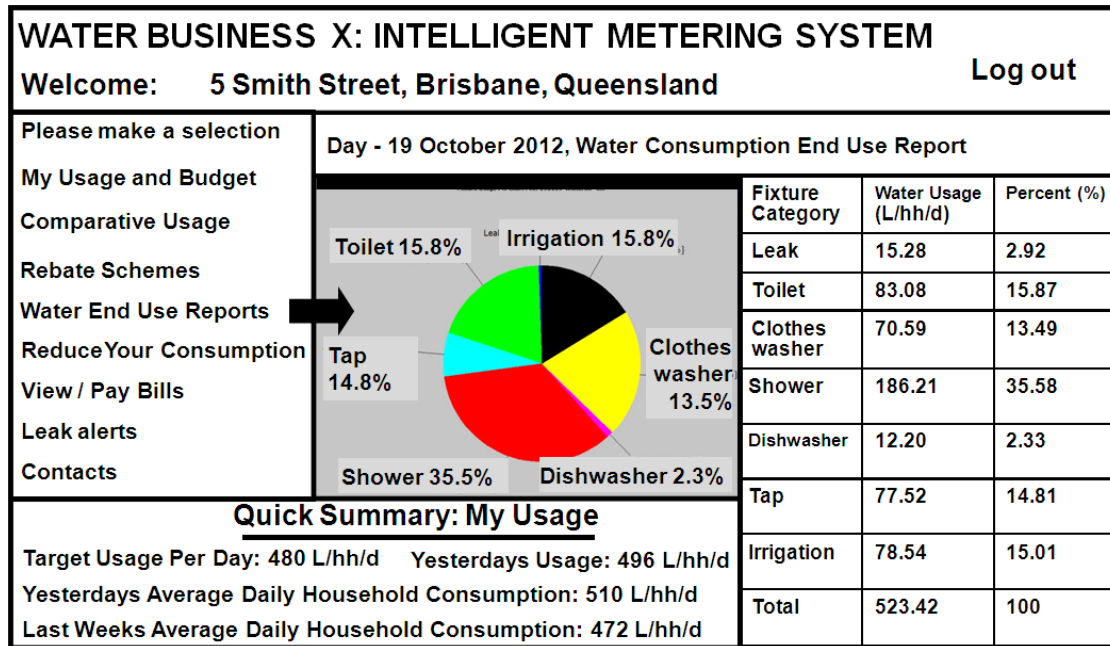
8.   Vision of an intelligent water metering and information system

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 14 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.

WATER BUSINESS X: INTELLIGENT METERING SYSTEM
Welcome: 5 Smith Street, Brisbane, Queensland   Log out

Please make a selection

My Usage and Budget

Comparative Usage

Rebate Schemes

Water End Use Reports

Reduce Your Consumption

View / Pay Bills

Leak alerts

Contacts

Day - 19 October 2012, Water Consumption End Use Report

Toilet 15.8%   Irrigation 15.8%

Tap 14.8%   Clothes washer 13.5%

Shower 35.5%   Dishwasher 2.3%

Fixture Category	Water Usage (L/hh/d)	Percent (%)
Leak	15.28	2.92
Toilet	83.08	15.87
Clothes washer	70.59	13.49
Shower	186.21	35.58
Dishwasher	12.20	2.33
Tap	77.52	14.81
Irrigation	78.54	15.01
Total	523.42	100

Quick Summary: My Usage

Target Usage Per Day: 480 L/hh/d   Yesterdays Usage: 496 L/hh/d

Yesterdays Average Daily Household Consumption: 510 L/hh/d

Last Weeks Average Daily Household Consumption: 472 L/hh/d

Figure 14 Illustrative customer interface for a proposed water management information system that employs the developed *Autoflow* firmware

9. Conclusion

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 feasible thanks to the development of *Autoflow*. The software tool offers a robust pattern recognition procedure through the hybrid combination of customised HMM, ANN and DTW algorithms. The verification process demonstrated that a considerable increase in recognition accuracy has been achieved when applying the new analytics methods underpinning Version 2 of *Autoflow*. This version achieved an overall water end use event pattern recognition accuracy (i.e. average for all end use categories) of just over 90%, which is near the target of 95% that is considered sufficient for a commercial application of the software.

The final analytical stage to be completed in future research is to further develop the existing self-learning function within *Autoflow* (i.e. create Version 3). New techniques are required which assist the existing self-learning function to better characterise the distinct end-use characteristics of any new residential house analysed by the system, by drawing on existing extensive registries of event signature prototypes already established and interpreting these new variant events and accurately assigning them to the prototype registry. Further training, testing and validation using samples from independent homes from various urban areas within Australia, and later abroad, will also be carried out to confirm the accuracy level of this application for various situational context (e.g. country, region, dwelling type, etc.). Ultimately, the goal is to create a highly accurate, adaptable and autonomous *Autoflow* firmware application that has widespread commercial application.

References

Baum, L. E. Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37 (6),1554–1563.

Baum, L. E. Petrie, T. Soules, G. Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics* 41 (5), 164 -178.

Beal, C., Stewart, R.A., Huang, T.T., Rey, E. (2011). SEQ residential end use study. *Journal of the Australian Water Association,* 38 (1), 80-84.

Beal, C.D. Stewart. R.A. (2014) Identifying residential water end uses underpinning peak day and peak hour demand. *Journal of Water Resources Planning and Management*, 140(7), 1268 – 1278.

Chang, F.C. & Huang, H.C. (2012) A refactoring method for cache-efficient swarm intelligence algorithms. *Information Sciences*, 192 (1), 39-49.

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

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

CSIRO., (2010). Water supplies could halve in two decades [Online]. Perth: Perthnow. Available: http://www.perthnow.com.au/news/western-australia/water-supplies-could-halve-in-two-decades/story-e6frg13u-1225836669765 [Accessed 26-10-2010].

Ephraim, Y., Merhav, N. (2002). Hidden Markov processes. *IEEE Transactions on Information Theory*, 48(6), 1518-1569.

Froehlich, J.E., Larson, E., et al., (2009). HydroSense: Infrastructure-mediated single-point sensing of whole-home water activity. *Prococeedings of UbiComp 2009*, Orlando, Florida, USA, 235–244.

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

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

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*, 319, 756-760.

Heaton, J. (2008) Introduction to Artificial Neural Network with Java, second edition. *Heaton Research Inc*. ISBN 1-60439-008-5.

Karayiannis, N. & Venetsanopoulos, A. (1993), Artificial Neural Networks, *Boston: Kluwer Academic Publishers.*

Latif, A. (2013) An adaptive digital image watermarking scheme using fuzzy logic and tabu search. *Journal of Information Hiding and Multimedia Signal Processing*, 4, 250-271.

Manmatha, R. and Srimal, N. (1999) Scale Space Technique for Word Segmentation in Handwritten Manuscripts. P*roceedings of 2nd Int'l Conf. on Scale-Space Theories in Computer Vision*, Corfu, Greece, September 26-27, 1999, pp. 22-33.

Manmatha, R. and Rath, T. M. (2002) Word Image Matching Using Dynamic Time Warping. *Multi-Media Indexing and Retrieval Group, Center for Intelligent Information Retrieval, University of Massachusetts*. Technical Report

Marquez, J.P. (2001) Pattern recognition: concepts, methods and applications. *Springer*, ISBN: 3-540-422978.

Muller, M. (2007). Information Retrieval for Music and Motion, Chapter 4. *Springer*, ISBN 978-3-540-74047-6.

Myers, C. S., Rabiner, L. R. (1981). A comparative study of several dynamic time-warping algorithms for connected word recognition. *The Bell System Technical Journal*, 60, 1389-1409.

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 Int. Ass. for Hydro-Environment Engineering and Research*, 2137-2144, Brisbane, Australia, 26 June-1 July, 2011.

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), 182-201.

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.

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.

Rabiner, L., Juang, B. (1993). Fundamentals of speech recognition. Prentice-Hall, Inc., Chapter 4. Rabiner, L. R. 1990. A tutorial on hidden Markov models and selected applications in speech recognition. Readings in speech recognition. *Morgan Kaufmann Publishers Inc*.

Sahin, O. Siems, R.S. Stewart, R.A. Porter, M.G. (2014a) Paradigm shift to enhanced water supply planning through augmented grids, scarcity pricing and adaptive factory water: A system dynamics approach. *Environmental Modelling & Software*, Available online 19 June 2014, ISSN 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2014.05.018.

Sahin, O. Stewart, R.A. Porter, M.G. (2014b) Water security through scarcity pricing and reverse osmosis: a system dynamics approach, *Journal of Cleaner Production*, 88, 160-171.

Sakoe, H., Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1), 43-49.

Satish, L., Gururaj, B. I. (1993). Use of hidden Markov models for partial discharge pattern classification. *IEEE Transactions on Electrical Insulation*, 28(2), 172-182.

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.

Stewart, R.A., Willis, R.M., Panuwatwanich, K. and Sahin, O. (2013). Showering behavioural response to alarming visual display monitors: longitudinal mixed method study. *Behaviour & Information Technology*, 32(7), 695-711.

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

Zhang, D.L., Qiao, J., Li, J.B., Qiao, L.Y., Chu, S.C. and Roddick, J.F. (2014) Optimizing matrix mapping with data dependent kernel for image classification. *Journal of Information Hiding and Multimedia Signal Processing*, 5 (1), 72-79.