Design of a Fuzzy Logic Based Framework for Comprehensive Anomaly Detection in Real-World Energy Consumption Data

Muriel Hol[(⊠)] and Aysenur Bilgin

Institute for Logic, Language and Computation, 1098 XG Amsterdam, Netherlands
Muriel.Hol@student.uva.nl, A.Bilgin@uva.nl

Abstract. Due to the rapid growth of energy consumption worldwide, it has become a necessity that the energy waste caused by buildings is explicated by the aid of automated systems that can identify anomalous behaviour. Comprehensible anomaly detection, however, is a challenging task considering the lack of annotated real-world data in addition to the real-world uncertainties such as changing weather conditions and varying building features. Fuzzy Logic enables modelling knowledge-based non-linear systems that can handle these uncertainties, and facilitates modelling human interpretable systems. This paper proposes a new method for annotating anomalies and a novel framework for interpretable anomaly detection in real-world gas consumption data belonging to the educational buildings of the Hogeschool van Amsterdam. The proposed architecture uses the Wang and Mendel rule learning with k-means clustering and does not require prior knowledge of the data, while preserving transparency of the model behaviour. Experiments have shown that the proposed system matches the performance of existing baselines using an artificial neural network while providing additional desired features such as transparency of the model behaviour and interpretability of the detected anomalies.

Keywords: Fuzzy logic systems · Energy consumption · Anomaly detection · Forecasting

1 Introduction

Population growth, improved indoor comfort and services, and increased time spent inside buildings cause a rapid worldwide growth of energy consumption [14]. Additionally, economic growth is causing a rise in the energy required for the buildings in the service sector such as schools, hospitals and recreational buildings. In Europe, buildings consume 40% of the entire energy, and non-residential buildings comprise the majority of this [12].

In a research conducted in 2013 [6], it was observed that buildings waste up to 30% of energy due to deficient management, which can be prevented by the utilisation of automated fault detection and diagnostics (FDD) [9]. FDD

The code is available at: https://github.com/murielhol/FuzzyEnergy.

systems detect abnormal behaviour and provide explicit information about the cause of the problem in order to enable targeted management. This detection of abnormal behaviour is described as anomaly detection [4].

The demand for better energy management in buildings through anomaly detection has resulted in various studies in the field of forecasting energy consumption [22]. The energy consumption of a building is influenced by complex features such as the buildings' materials, the users' schedule, the weather and the occupants' subjectivity regarding indoor comfort. Therefore, forecasting energy consumption requires algorithms that handle non-linearity and uncertainty.

Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are the most frequently used algorithms for forecasting energy consumption as they perform well with non-linear data [7,22]. Both methods are referred to as black-box algorithms, which means that there is no natural language explanation of the models' behaviour and hence no interpretation of mapping the input to the output of the systems [15]. As previously stated, information about the anomaly is valuable for improved energy management.

Similar to the forecasting of energy consumption, the detection of anomalies in energy data requires techniques that can handle uncertain and non-linear data and can enable transparency of the process [20]. Fuzzy Logic (FL) satisfies these requirements. After the introduction of FL by Lotfi A. Zadeh in 1965, FL has achieved success in a wide range of research fields such as control systems, image processing, industrial automation, robotics, and optimisation [17]. Zadeh states that: "Essentially, such a framework provides a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership rather than the presence of random variables" [21] (Zadeh, p. 339). Furthermore, FL uses linguistic terms, which facilitate understanding of the model behaviour.

The applications of FL in forecasting energy consumption and anomaly detection have yielded good results in the previous studies [18,20]. Rocha et al. [15] strengthen the argumentation for exploring FL by stating that although complex algorithms such as ANN and SVM may lead to high accuracies, non-complex transparent algorithms such as FL can provide valuable insights regarding the model behaviour.

For buildings in the service sector of The Netherlands, gas consumption comprises the largest share in energy consumption [13]. Gas consumption is severely affected by the weather, which is an uncertain feature. Therefore, the aim of this study is to tackle the uncertainties inherent to forecasting and anomaly detection in gas consumption data using FL and to provide linguistic descriptions of the identified anomalies in order to facilitate improved energy management systems. The main contributions reported in this paper are the introduction of a new method for annotating anomalies, and a novel framework for anomaly detection based on FL.

The paper is structured as follows: In Sect. 2, we present a brief background of the methods used in this research. Section 3 focuses on the related work. Section 4 details the proposed framework while Sect. 5 presents the experiments and results. In Sect. 6, we discuss our findings and lastly, we provide conclusions and future work in Sect. 7.

2   Background

This section provides a brief theoretical background on FL, Fuzzy Logic Systems, Anomaly Detection and supplementary methods we have used.

2.1   Fuzzy Logic

Zadeh describes a fuzzy set as "a class with continuum grades of membership" [21]. In other words, an element can belong to a set with a degree of membership, as opposed to classical logic, which only allows binary membership. For example, a linguistic variable, say '*Temperature*', can be modelled using two linguistic terms: '*Cold*' and '*Warm*'. These linguistic terms are characterised by membership functions, which are denoted $\mu_{Cold}(\boldsymbol{x})$ and $\mu_{Warm}(\boldsymbol{x})$, respectively. The membership functions (MFs) are used to assign membership degrees to $\boldsymbol{x}$ within the unit interval $[0, 1]$. Figure 1a illustrates the overlapping of MFs where both $\mu_{cold}(\boldsymbol{x}) > 0$ and $\mu_{warm}(\boldsymbol{x}) > 0$ can hold for the same value of $\boldsymbol{x}$. However, in Fig. 1b, the Boolean sets only allow for either $\mu_{cold}(\boldsymbol{x}) = 1$ or $\mu_{warm}(\boldsymbol{x}) = 1$.

Modelling of MFs can be regarded as a highly problem-dependant task. Common techniques include using expert knowledge and using training data to which the MFs can be fitted. The hyper parameters to be taken into account are the number of fuzzy sets per feature, their domain (universe of discourse) and the shape of the MFs. An important design decision is the amount of overlap between the MFs. The most frequently used shapes for MFs include Triangles, Trapezoids and Gaussians.
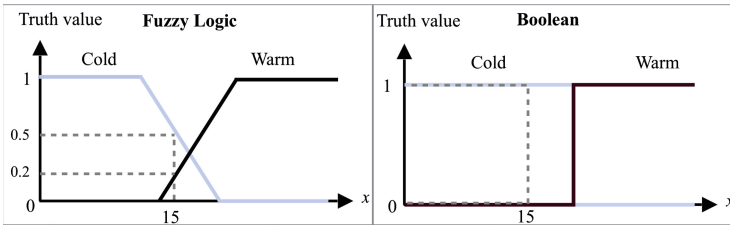

Fig. 1. (Left) a: FL example, $\mu_{Cold}(15) = 0.5$ and $\mu_{Warm}(15) = 0.2$. (Right) b: Boolean example $\mu_{Cold}(15) = 1$ and $\mu_{Warm}(15) = 0$.

2.2   Fuzzy Logic Systems

A Fuzzy Logic System (FLS), also named as Fuzzy Inference System (FIS) or Fuzzy Rule-Based System (FRBS), consists of 5 components: (1) Fuzzifier, (2) Rule Base, (3) Data Base, (4) Inference Engine and (5) Defuzzifier [16]. Figure 2 provides an overview of the FLS architecture. FLSs are universal approximators [19], capable of learning any non-linear function.

– **Fuzzification:** Fuzzification is the process of assigning a membership degree to a crisp input for a linguistic term. For example, in Fig. 1a, the fuzzification of the crisp input $x = 15$ yields 0.5 for $\mu_{cold}(15)$ and 0.2 for $\mu_{warm}(15)$, as marked on the figure.
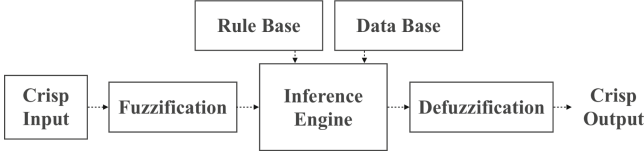
Fig. 2. Fuzzy logic system architecture

– **Rule Base:** The allowance of uncertainty reflects the human reasoning process and ambiguity of the discourse [2]. This reasoning is captured in the format of fuzzy IF-THEN Rules. The IF part of the fuzzy rule holds the antecedents and the THEN part holds the consequents. These rules can be given by an expert, or can be learnt from training examples [2]. A fuzzy rule can be formalised as:

$$\text{IF } x_1 \text{ is } \mathcal{A}_1 \text{ and ... and } x_p \text{ is } \mathcal{A}_p \text{ THEN y is } \mathcal{B} \tag{1}$$

– **Data Base:** The data base stores the linguistic variables (e.g. temperature), their linguistic terms (e.g. cold, warm) and the parameters of the MFs.
– **Inference Engine:** The inference engine infers the firing strengths of the rules for the crisp inputs, using the information from the data base and the rule base. In order to calculate the firing strength of a rule (where the antecedents are connected using the logical operator AND), the fuzzy inter-section[1] (i.e. t-norm) is used [11]. The most commonly used operators for t-norm are the minimum and the product. Using the minimum operator, the firing strength of a rule where there are p antecedents can be calculated as follows:

$$FS = min(\mu_{\mathcal{A}_{x_1}}(x_1), \mu_{\mathcal{A}_{x_2}}(x_2)...., \mu_{\mathcal{A}_{x_p}}(x_p)) \tag{2}$$

In a Mamdani type inference, both the antecedents and the consequents are fuzzy sets, which leads the output of the inference engine to be a fuzzy set.
– **Defuzzification:** Defuzzification is used to convert the fuzzy output of the inference engine into a crisp output. Among several defuzzification methods, Eq. 3 formalises the centroid defuzzification where K is the number of rules, fs_i is the firing strength of the ith rule, and c_i is the centroid of the consequent of the ith rule:

$$y = \frac{\sum_{i=1}^{K} fs_i \cdot c_i}{\sum_{i=1}^{K} fs_i} \tag{3}$$

2.3   Anomaly Detection

Anomaly detection is the detection of abnormal behaviour of a data point. A frequently used method for anomaly detection is forecasting using time-series data. A data point is classified as an anomaly when the squared difference between the

---

[1] In FL, set theoretic operations (e.g. intersection, union) are defined in terms of their membership functions.

predicted and actual value exceeds a predefined threshold [4]. In their extensive survey on anomaly detection, Chandola and Kumar [4] state that: "Defining a normal region that encompasses every possible normal behaviour is very difficult. In addition, the boundary between normal and anomalous behaviour is often not precise. Thus, an anomalous observation that lies close to the boundary can actually be normal, and vice versa." (p. 3). Accordingly, it is very challenging to measure the classification performance (i.e. normal vs. anomalous) using (unannotated, real-world) data that has uncertainties with regards to what is considered as normal or anomalous.

2.4   Supplementary Methods Used in the Proposed Framework

The proposed framework uses the Wang and Mendel (WM) rule learning with k-means clustering. In this section, both methods will be presented, together with the performance measures we employed for the evaluation.

Wang and Mendel Rule Learning. Wang and Mendel provide a simple adhoc model for learning a fuzzy rule base from data. Their method is renowned for its "simplicity and good performance" [2] (Alcala et al. p. 11) and is referred to as the WM method. The method is based on the following steps [2]:

1. For each training example $e_l = \{x_1^l, x_2^l..., x_n^l, y^l\}$, find the sets $[\mathcal{A}_q^1...\mathcal{A}_r^n, \mathcal{C}_s]$ that e_l has the highest membership to, and create the rule

$$R_j = \text{ IF } x_1 \text{ is } A_q^1 \text{ and... and } x_n \text{ is } A_r^n \text{ THEN } y \text{ is } C_s \qquad (4)$$

   with degree $D_j = \mu_{\mathcal{A}_q}(x_1^l) \cdot ... \cdot \mu_{\mathcal{A}_r}(x_n^l) \cdot \mu_{\mathcal{C}_s}(y^l)$.
2. In order to prevent conflicting rules, if the rule base already has a rule that has the same antecedents with a different consequent, only keep the rule that has the highest degree D_j.

K-Means. K-means is a clustering algorithm that provides the centroids of the clusters in the data. In the proposed approach, the cluster centroids will be assigned as the centroids of the fuzzy sets.[2]

Performance Measures. In order to measure the forecasting performance, we employed the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). For the anomaly detection performance, we employed the F1 score, which is a classification measure calculated with the precision and recall.

3   Related Work and Motivation

This research was mainly inspired from de Nadai and van Someren [7], who employed an ANN for the forecasting of gas consumption data. They further

---

[2] sklearn.cluster.KMeans: http://scikit-learn.org/stable/modules/clustering.html#clustering. The default settings were used in this paper.

improved their approach using an auto-regressive integrated moving average model (ARIMA) [7], which models seasonal data. The ARIMA predicted gas consumption is then added as a feature to the ANN model. Using the ARIMA features, their initial MAE of the ANN was improved from 9.52 to 7.33.

Zhao and Magoulés [22] and Ahmad et al. [1] provide an extensive review of research in energy consumption forecasting. Both studies address that SVM and ANN are the most frequently used approaches. Additionally, the studies both confirm the improvement achieved through the utilisation of ARIMA models. Neither of the studies discuss FL as an individual approach, however, they both mention several studies that obtained improved results when FL was integrated in order to form a hybrid model.

On the other hand, while considerable research is done on the prediction of energy consumption [22], few equivalent research was found on anomaly detection. One reason for this is the absence of annotated data and its consequential challenges for the performance evaluation. The lack of annotated data can be dealt with by the injection of synthetically created anomalous data [20]. However, when few and obviously anomalous synthetic anomalies are used for the evaluation, the reliability of the models become questionable. It can also be argued that the boundary problem is ignored, and hence, the models that use synthetic anomalies fail to reflect the system performance accurately, which is crucial especially in real-world scenarios.

Additionally, anomaly detection with supervised learning requires annotated data in order to confirm that the training data is free from anomalies. Moreover, the assumption that the training data is clean "constitutes a fundamental concept underlying the use of anomaly detection techniques" [20] (Wijayasekera et al. p. 1832). However, labelling the data as anomalous or normal is an expensive task that is mostly done manually. Hodge and Auston [8] provide a framework for anomaly detection. They provide supervised methods that can be used to "identify errors and remove their contaminating effect on the data set and as such to purify the data for processing" (p. 1).

An important characteristic of the anomaly detection task is the format in which the anomaly is reported [4]. Linguistic output about an anomaly allows understanding of the cause of the problem and therefore facilitates targeted action. FL naturally provides the information about the reasoning process of the system. Wijayasekera et al. successfully implemented a fuzzy anomaly detection method that avails from the linguistic reasoning of the model [20]. They propose to make the provided information about the anomaly more concise by only displaying the most important fired antecedents. They show that the firing strengths and linguistic terms of the sets can be used not only to provide information about the anomaly, but also to adjust the content of the feedback to be communicated back to the user. In our proposed method, we adopt this idea of information summarising to enable comprehensive anomaly detection.

When the Rule Base is learnt in an ad-hoc manner, it represents the behaviour of the training data [2]. Hence, when the training data represents the normal behaviour, the firing strength of a data point indicates to what extent the data point behaves normally [20]. The proposed approach exploits this feature by

using the firing strengths for the anomaly classification, instead of the predicted crisp output.

4 Proposed Framework for Anomaly Detection in Gas Consumption Data

This section presents the data and the preprocessing steps, which are followed by the design of the proposed approach.

4.1 Energy Consumption Data and Feature Extraction

In this paper, we used the gas and electricity consumption rates of The Nicolaes Tulphuis (NTH) building, which belongs to HvA[3]. This data originates from

Table 1. The features used in the proposed approach

Feature	Data	Domain
Radiation	Q(t) (J/cm^2)	[0, 334]
Temperature	T(t) (0.1 C)	[−176, 331]
Wind speed	FH(t) (0.1 m/s)	[0, 220]
Humidity	U(t) (%)	[22, 100]
Temperature peak in past 5 h	$\mathrm{argmax}_{x\in[t-6,t-1]} T(x)$	[−176, 331]
Temperature difference with previous hour	$T(t) - T(t-1)$	[−57, 63]
Gas consumption (target)	G(t)	[0, 323]
Gas consumption 1 h before	G(t–1)	[0, 323]
Gas consumption 2 h before	G(t–2)	[0, 323]
Gas peak in past 5 h	$\mathrm{argmax}_{x\in[t-6,t-1]} G(x)$	[0, 323]
Gas peak in past 24 h	$\mathrm{argmax}_{x\in[t-25,t-1]} G(x)$	[0, 323]
Gas sum of past 5 h	$\sum_{i=1}^{5}$ G(t-i)	[0, 1098]
Gas sum of past 24 h	$\sum_{i=1}^{24}$ G(t-i)	[0, 4116]
Gas mean of past 15 days	$\frac{1}{360}\sum_{i=1}^{5}$ G(t–361)	[0.41, 132.11]
Day of the week		[0, 6]
Next day of the week		[0, 6]
Hour of the day		[1, 24]
Day of the year		[0, 366]
Electricity consumption (in kWh)	E(t)	[0, 53.11]
Electricity consumption peak in past 5 h	$\mathrm{argmax}_{x\in[t-6,t-1]} T(x)$	[0, 53.33]
STD residuals for trend = day	resDay(t)	[−116.24, 109.28]
STD residuals for trend = year	resYear(t)	[−117.85, 164.19]

---

[3] Hogeschool van Amsterdam: http://www.hva.nl/over-de-hva/locaties/locaties.html.

the research in de Nadai and van Someren [7] and consists of 52608 hourly gas consumption timestamps ranging from the first day of 2008 until the first day of 2014. There were 2 missing dates, 29 missing gas values and 6 missing electricity values, which were linearly interpolated.

We adopted the features that are used by de Nadai and van Someren [7] in order to enable a fair comparison with their results. Moreover, the same features were also adopted by Lodewegen [10], who also studied the performance of ANN on this data set. Both studies will serve as reference points for the results presented in this paper. In total, there are 22 features (see Table 1), which can be categorised under the following 4 major interests:

1. **Weather data:** Hourly weather data was obtained from the KNMI website[4]. The Schiphol weather station is the closest one to the NTH building (± 20 km).
2. **Energy consumption time series:** Each data row has extensive information about the historical gas and electricity consumption, such as the gas consumption in the previous hour or the peak in consumption during the last 24 h.
3. **Time stamps:** Information about the moment when the data was recorded is provided in terms of day of the week and the hour of the day. The next day of the week is also included as a feature, since schools tend to warm up the buildings for Monday in advance [7]. National holidays were labelled as Sundays.
4. **Seasonal Trend Decomposition:** Gas consumption is strongly related to seasonal trends in for example days and years. In order to allow the model to understand the behaviour of the consumption in a disconnected manner from its seasonal trend, the LOESS method [5] for Seasonal Trend Decomposition (STD) is adopted [7]. By using the LOESS method, the gas consumption is decomposed into seasonality, trend and residual. As emphasised by [7], the residual is very informative for the prediction of gas consumption. Figure 3 shows the decomposing for the days and years from the beginning of 2008 until the end of 2013.
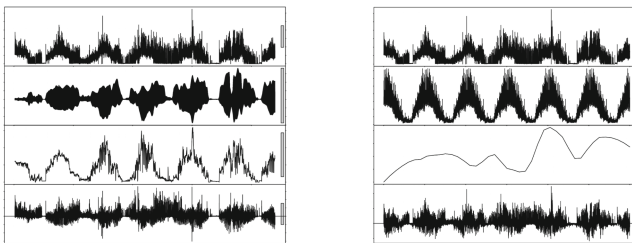

Fig. 3. Seasonal trend decomposition of the gas consumption. From top to bottom: the data, the seasonal component, the trend and the residuals. (Left) a: STD for trend day (frequency is 24). (Right) b: STD for trend year (frequency is 24*365.5).

---

[4] Koninklijk Nederlands Meteorologisch Instituut: http://projects.knmi.nl/klimatolog ie/uurgegevens/selectie.cgi.

In the proposed FLS, each feature is a linguistic variable such as *Temperature*, and the feature values are linguistic terms such as *very low, high, somewhat high* etc.

4.2   Pre-processing of the Data and Annotation of Anomalies

In order to detect anomalies, the data needs to be labelled as normal or anomalous. Since the original data is not annotated, we propose to synthetically label the data using the 3 sigma rule and the Mahalanobis distance. The 3 sigma rule assumes that the probability of a data point lying in the region ± 3 times the standard deviation (σ) of the mean (μ) is 0.997. This is formalised as follows: $P(\mu - 3\sigma \leq X \geq \mu + 3\sigma) \approx 0.997$.

Since the data is multivariate, the Mahalanobis distance measure is used to calculate the distance to the mean vector [8]. Mahalanobis distance measure is formalised as follows: $MD = \sqrt{(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})}$. Here, Σ is the covariance matrix. The covariance matrix is a matrix that holds the variance for each feature combination. Therefore, the dependencies between all features are taken into account. In the rest of the paper, the inliers will be referred to as 'clean data'. We employ k-fold cross validation where $k = 5$, hence, the data will be split into 5-folds of 80% for training, and 20% for testing.

4.3   Architecture of the Proposed Framework

The proposed framework is illustrated in Fig. 4 and is based on the following steps:

1. Data Base construction: We used k-means clustering to find an optimal centroid location for each MF. The linguistic terms that describe the MFs depend on the number of clusters and are listed below:
   – For 11 clusters: Extremely low, very low, low, somewhat low, a bit low, medium, a bit high, somewhat high, high, very high, extremely high
   – For 9 clusters: Extremely low, very low, low, a bit low, medium, a bit high, high, very high, extremely high
   – For 7 clusters: Extremely low, very low, low, medium, high, very high, extremely high
2. Rule Base construction: We adopted the ad-hoc rule learning method constructed by Wang and Mendel [19] as presented in Sect. 2.4.
3. Anomaly classification threshold: A threshold was constructed based on the firing strengths of the anomalies using the part of the data between 2008 and 2012. For a chosen anomalous data point, we calculated the mean firing strength using the entire rule base. We repeated this for all the anomalous points that are obtained from the 3 sigma rule. We then calculated the mean firing strength of all 480 anomalous points and used this as a threshold ($\overline{f}$). A data point is classified as an anomaly when its mean firing strength over all the rules is lower than $\overline{f}$.
4. Other FLS design parameters: We used Mamdani inference, with minimum t-norm and centroid defuzzification.
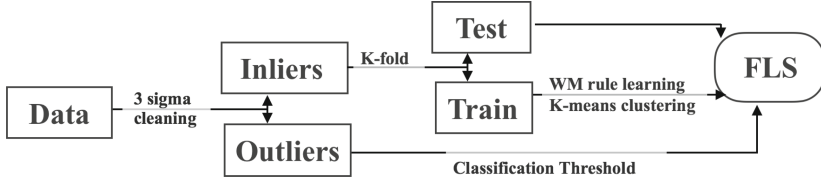
Fig. 4. The architecture of the proposed framework

5 Experiments and Results

In this section, we present several experiments on the following: (1) tuning of the FLS, (2) validation of the forecasting performance and (3) validation of the anomaly detection performance. Furthermore, we illustrate the capabilities of FL based approach by describing a couple of anomalies.

5.1 Tuning Experiments and Results

The tuning of the FLS consists of several cycles of parameter tuning until the lowest forecasting error is obtained through trial and error. Once the configuration is found, then the final rule base and data base are stored. The tuning tests were done on the data from the year 2013, in order to reflect real life performance in which the system is trained on historical data and tested on the forthcoming.

The system is tuned for (a) the MF type (i.e. Triangle, Trapezoidal or Gaussian) (b) the number of MFs per feature, and (c) the domains of the MFs. For tuning the MF type, we fixed the number of MFs to be 9 per feature. The initial variance of the Gaussian MF was 0.1 (in a normalised domain of 0 to 1) and the initial base of the Triangle/Trapezoidal MF was 0.2 (in a normalised domain of 0 to 1). Gaussian MFs gave the best performance.

Table 2. The number of clusters (k) per feature (after tuning)

k	Features
11	Radiation, Temperature, Wind speed, Humidity, Gas consumption (target), Gas sum of past 24 h, Hour of the day, Day of the year
9	Temperature peak in past 5 h, Gas peak in past 24 h, Gas sum of past 5 h, Gas mean of past 15 days, Electricity (in kWh), Electricity peak in past 5 h, STD residuals for trend = day, STD residuals for trend = year
7	Gas consumption 1 h before, Gas consumption 2 h before, Gas peak in past 5 h, Day of the week, Day of the year

For tuning the number of MFs per feature (i.e. linguistic variable), we have chosen to place the mean of the Gaussian MFs to be the cluster centre obtained from k-means clustering that was conducted on the mean target gas values of each training example. Formally, for each unique feature value x, we calculate
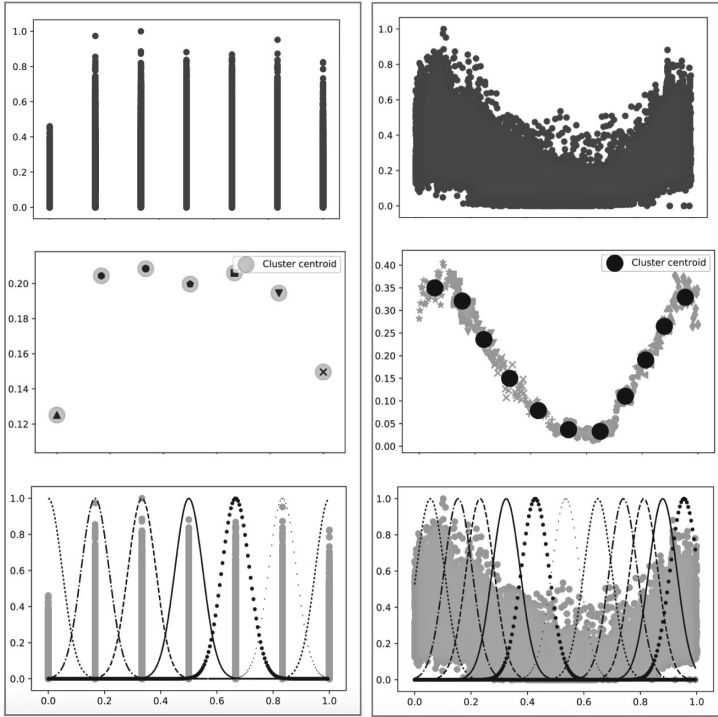
Fig. 5. The process of tuning the features: (Left) day of the week and (Right) day of the year. From top to bottom: (1) The data scaled between 0 and 1, (2) the cluster centroids according to k-means on the mean target values (3) fitting the MFs using their mean at the centre of the clusters, and a variance of 0.07.

the mean gas consumption and collect these in a vector that is as large as the number of unique feature values $\overline{y}$. E.g. if the data contains 3 examples which 'Temperature' value is 5, and they have a 'Gas consumption' value of 1, 2 and 3, respectively, then the $\overline{y}_{temperature=5} = 2$. Then, we perform k-means on x and $\overline{y}$ with a predefined $k \in 5, 7, 9, 11$. These values were chosen by trial and error until the MAE did not decrease. In order to prevent over-fitting, the maximum k was set to 11. For some features, the number of clusters were manually assigned based on the authors' visual observations on data plots. The number of clusters per feature is displayed in Table 2.

After having learnt the cluster centres, which are to be used as the mean of the Gaussian MFs, the variance of the Gaussian MFs was tuned using the following values for variance $\sigma \in \{0.06, 0.07, 0.08, 0.09, 0.1, 0.2\}$. Configuration using $\sigma = 0.07$ gave the lowest forecasting error (i.e. MAE). The values were chosen by trial and error. Figure 5 demonstrates the process of tuning for two different features.

Table 3. 5-fold validation results: (Left) The results on the clean data set. (Right) The results on the full data set

Fold	RMSE	MAE	F1-score	Number of rules	RMSE	MAE	Number of rules
1	10.11	7.76	0.53	40906	14.01	11.09	41145
2	10.15	7.8	0.56	40814	17.47	13.39	41431
3	10.47	7.85	0.53	41035	14.32	10.78	41406
4	10.92	8.40	0.58	40821	14.72	11.28	41249
5	12.90	9.84	0.40	40858	17.14	13.28	40291
Mean	10.91	8.33	0.539	40886.8	15.53	11.96	41104
STDEV	1.035	0.788	0.033	80.982	1.624	1.265	469.5

5.2   Forecasting Experiments and Results

By training the system (using WM rule learning) on the clean data, we modelled the 'normal' behaviour of the gas consumption in the NTH Building. We evaluated the model for the forecasting performance using the RMSE and the MAE. The results of 5-fold cross validation on the cleaned data are listed in Table 3. Since the results of the referenced ANN approaches of de Nadai and van Someren [7] and Lodewegen [10] are reported on the full data set, for a fair comparison, we provide the forecasting results on the full data set, as well. The results on the cleaned data set yield an RMSE of 10.91 and an MAE of 8.33. For the full data set, the MAE is 15.53 and the RMSE is 11.96.

We performed statistical T-tests in order to compare the results of the FLS and the ANN approaches. We observed that the MAE and RMSE of the proposed FL based framework are significantly higher than the ANN approach proposed by de Nadai and van Someren [7], however, not significantly different when compared to the approach of Lodewegen [10]. Table 4 shows the comparison results of all three approaches.

Table 4. Comparison with ANN approaches [7,10] on the full data set

Approach	RMSE	MAE	Configuration
de Nadai and van Someren [7]	N/A	9.52	15 epochs, 60 neurons, 1 hidden layer
Lodewegen [10]	15.58	12.01	150 epochs, 150 neurons, 1 hidden layer
Proposed FL Approach	15.54	12.07	Gaussian mf, 7–11 sets per feature, 41104 rules

5.3   Anomaly Detection Experiments and Results

In order to evaluate the performance of the anomaly detection, we annotated the data as described in Sect. 4.2. Our approach for anomaly detection relies on the significant difference between the firing strengths of the inliers and the outliers (i.e. anomalous data). For the purpose of illustrating that the difference is indeed significant, we plotted the firing strengths of both categories in Fig. 6. The mean firing strength of the inliers is 5.33e–05 (Fig. 6a) whereas the mean firing strength of the outliers is 5.79e–07 (Fig. 6b). Hence, it can be deduced that the difference between the mean firing strength of inliers and the outliers is significant. The performance on anomaly detection is validated using the same cross-validation folds as we used for the forecasting. The F1-score results for each fold are listed in Table 3.
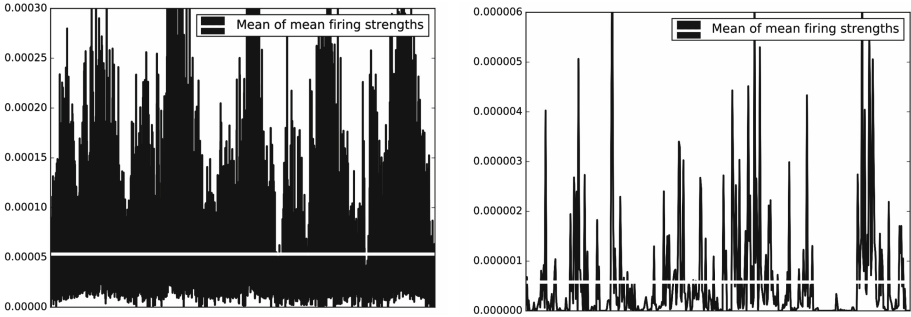

Fig. 6. (Left) a: the average firing strength of the 52128 inliers with a mean average of 5.33e–05. (Right) b: The firing strengths of the 480 outliers with a mean average of and 5.79e–07.

For further evaluation and comparison purposes, a baseline for anomaly detection has been set. A classifier that classifies an anomaly with 50% chance has been chosen to be the baseline. For the baseline system, we obtained an F1-score of 0.074. For the proposed framework, we observed the average F1-score over 5-folds to be 0.539, which is far above the baseline. However, the imbalance of the classes within data, which consists of 52129 inliers and 481 outliers, has an influence on the F1-score. Therefore, the confusion matrix with the True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) is presented in Fig. 7 in order to provide full information on the classification performance.

5.4   Linguistic Description of Anomalies

In order to exploit the advantages of the proposed FL based framework and to allow for interpretability of the anomalies, we adopted the method used by Wijayasekera et al. [20]. Accordingly, the anomalies can be described using the fuzzy rules. Furthermore, the most influential features can be detected by looking at the least fired antecedents of the rule. To give an example, one of the data

Classified

	p	n
p′	383 **0.798**%	97 0.201%
n′	568 0.054%	9859 **0.946**%

Actual value

```
HUMIDITY IS A BIT HIGH
ENERGY CONSUMPTION PEAK OF PAST 5 HOURS (KWH) IS VERY LOW
GAS CONSUMPTION PEAK IN THE PAST 5 HOURS IS HIGH
GAS CONSUMPTION SUM IN THE PAST 5 HOURS LOW
TEMPERATURE IS A BIT HIGH
TEMPERATURE DIFFERENCE WITH THE PAST HOURS IS SOMEWHAT LOW
RADIATION IS EXTREMELY LOW
```

```
DAY OF THE YEAR IS EXTREMELY LOW
ENERGY CONSUMPTION PEAK OF PAST 5 HOURS (KWH) IS LOW
GAS CONSUMPTION PEAK IN THE PAST 5 HOURS IS LOW
GAS CONSUMPTION ONE HOUR BEFORE IS MEDIUM
HOUR OF THE DAY IS A BIT LOW
DEVIATION OF THE DAILY GAS CONSUMPTION TREND IS EXTREMELY HIGH
GAS CONSUMPTION PEAK IN THE PAST 24 HOURS IS LOW
```

Fig. 7. Confusion matrix: left upper: TP, left bottom: FN, right upper: FP, right bottom: TN.

Fig. 8. Two examples of linguistic description of anomalies (Top) a: conflicting weather and (Bottom) b: 1st of January (holiday)

points that was classified as an anomaly is displayed with its 7 most influential features in Fig. 8a using the linguistic terms provided in Sect. 4.3. This example has low mean consumption in the past 5 h, however high peak in the past 5 h. This could be due to the combination of extremely low radiation yet a bit high temperature, and a bit high humidity (e.g. rain on a warm day may have caused anomalous behaviour). Another example is given in Fig. 8b. This was a data point recorded on the 1st of January, which was a holiday, and the seasonal residual for the year trend was extremely high. The linguistic descriptions can be very useful to the building managers, who wish to understand the cause of the anomalies and take targeted action.

6   Discussion

Although the forecasting results validate the performance of the FLS, it is important to note that the STD residuals at a time unit are used to classify a data point at that same time unit. Hence, if the system is only to be utilised for forecasting, then it would be better to adjust the features to include solely historical values (e.g. residuals at t–1). However, since an anomaly can be reported one time unit later, these are realistic features for anomaly detection.

With 22 features and more than 41700 training examples, the WM method for rule learning leads to a very large rule base. A rule reduction method, which is the Cooperative Rule Approach of Casillas et al. [3], was taken into consideration. However, it was found that the complexity of this method with N rules and k values for the target variable is at approximately $(N/k)^k$. Hence, for this research this would be $(10422/9)^9$, and therefore computationally infeasible. With regards to the parameter tuning, the order in which the parameters were tuned was arbitrarily chosen, which could therefore have lead to a local optimum.

7   Conclusions and Future Work

In this paper, we proposed a FL based framework for comprehensive anomaly detection in gas consumption data. The WM method and k-means clustering algorithm were combined into a supervised method for learning a fuzzy rule base, which represents the normal gas consumption behaviour of the NTH building. Furthermore, we introduced a new method for annotating anomalies using 3 sigma rule and Mahalanobis distance.

Regarding the forecasting of energy consumption, the performance of the proposed framework meets one of the existing approaches that was based on an ANN, however, is outperformed by the other. For the anomaly detection performance, we employed two techniques for evaluation: (1) we visually validated the efficiency of the proposed framework and (2) we compared the performance of the proposed framework with a baseline. We showed that the FL based approach is capable of detecting anomalies in gas consumption data far above the baseline. Furthermore, we exploited the advantages of FL based approach and demonstrated that the causes of the anomalies can be linguistically described.

For additional future work, we will investigate rule reduction techniques on high dimensional data. These techniques use optimisation methods such as simulated annealing. Finally, the informative capabilities of the proposed FL approach would be especially beneficial when applied on a data set that includes sensor data.

References

1. Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H., Saidur, R.: A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew. Sustainable Energy Rev. **33**, 102–109 (2014). doi:10.1016/j.rser.2014.01.069
2. Alcala, R., Casillas, J., Cordón, O., Herrera, F., Zwir, S.J.T.: Techniques for learning and tuning fuzzy rule-based systems for linguistic modeling and their application. In: Leondes, C. (ed.) Knowledge Engineering: Systems, Techniques and Applications, pp. 889–941. Academic Press, San Diego (1999)
3. Casillas, J., Cordón, O., Herrera, F.: Improving the Wang and Mendel's fuzzy rule learning method by inducing cooperation among rules. In: Proceedings of the 8th Information Processing and Management of Uncertainty in Knowledge-Based Systems Conference, pp. 1682–1688 (2000)
4. Chandola, V., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. **51**, 58 (2009). doi:10.1145/1541880.1541882
5. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: STL: a seasonal-Trend decomposition procedure based on loess. J. Official Statist. **6**, 3–73 (1990)
6. Colmenar-Santos, A., de Lober, L.N.T., Borge-Diez, D., Castro-Gil, M.: Solutions to reduce energy consumption in the management of large buildings. Energy Build. **56**, 66–77 (2013). doi:10.1016/j.enbuild.2012.10.004

7. De Nadai, M., van Someren, M.: Short-term anomaly detection in gas consumption through arima and artificial neural network forecast. In: Proceedings of 2015 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS 2015), pp. 250–255. IEEE, New York (2015). doi:10.1109/EESMS.2015.7175886

8. Hodge, V.J., Auston, J.: A survey of outlier detection methodologies. Artif. Intell. Rev. **22**, 85–126 (2004). doi:10.1007/s10462-004-4304-y. Kluwer Academic Publishers, Department of Computer Science, University of York

9. Katipamula, S., Brambley, M.R.: Review article: methods for fault detection, diagnostics, and prognostics for building systems: a review, Part I. HVAC & R. Res. **11**, 3–25 (2005)

10. Lodewegen, J.: Saving energy in buildings using an Artificial Neural Network for outlier detection. B.Sc Thesis. University of Amsterdam (2015). https://esc.fnwi.uva.nl/thesis/centraal/files/f1992667963.pdf

11. Mendel, J.M.: Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Direction. Prentice Hall, Upper Saddle River (2001). ISBN: 978-0130409690

12. ODYSEE: an analysis based on the ODYSSEE and MURE databases. http://www.odyssee-mure.eu/publications/br/energy-efficiency-trends-policies-buildings.pdf

13. ODYSEE. Energy efficiency trends and policies in the Netherlands (2015). http://www.odyssee-mure.eu/publications/national-reports/energy-efficiency-netherlands.pdf

14. Perez-Lombard, L., Ortiz, J., Pout, C.: A review on buildings energy consumption information. Energy Build. **40**, 394–398 (2008). doi:10.1016/j.enbuild.2007.03.007. Elsevier Inc

15. Rocha, A., Papa, J.P., Meira, L.A.A.: How far you can get using machine learning black-boxes. In: Conference on Graphics, Patterns and Images, vol. 16, pp. 1530–01834. IEEE (2010). doi:10.1109/SIBGRAPI.2010.34

16. Sanz, F., Ramrez, J., Correa, R.: Fuzzy inference systems applied to the analysis of vibrations in electrical machines. Fuzzy Inference Syst. Theory Appl. (2012). doi:10.5772/37448

17. Singh, H., Gupta, M.M., Meitzler, T., Hou, Z., Garg, K.K., Solo, A.M.G., Zadeh, L.A.: Real-Life Applications of Fuzzy Logic, Advances in Fuzzy Systems. Hindawi Publishing Corporation (2013). doi:10.1155/2013/581879

18. Suganthi, L., Iniyan, S., Samuel, A.A.: Applications of fuzzy logic in renewable energy systems: a review. Renew. Sustain. Energy Rev. **48**, 585–607 (2015). doi:10.1016/j.rser.2015.04.037

19. Wang, L.X., Mendel, J.M.: Generating fuzzy rules by learning from examples. IEE Trans. Syst. Syst. Man Cybern. **22**, 1414–1427 (1992)

20. Wijayasekera, D., Linda, O., Manic, M., Rieger, C.: Mining building energy management system data using fuzzy anomaly detection and linguistic descriptions. IEEE Trans. Industr. Inform. **10**, 1829–1839 (2014). doi:10.1109/TII.2014.2328291

21. Zadeh, L.A.: Fuzzy sets. Inf. Control **9**, 338–353 (1965)

22. Zhao, H., Magoules, F.: A review on the prediction of building energy consumption. Renew. Sustain. Energy Rev. **16**, 3586–3592 (2012). doi:10.1016/j.rser.2012.02.049