
Fan Zhang,1,2 Chris Bales,2 and Hasan Fleyeh1

1Department of Microdata Analysis, Dalarna University, Falun 79188, Sweden
2Department of Energy Technology, Dalarna University, Falun 79188, Sweden

Correspondence should be addressed to Fan Zhang; fzh@du.se

Received 13 July 2020; Revised 9 January 2021; Accepted 19 February 2021; Published 5 March 2021

Abstract

District heating systems that distribute heat through pipelines to residential and commercial buildings have been widely used in Northern Europe, and according to the latest study, district heating shares the most heat supply market in Sweden. Therefore, energy efficiency of district heating systems is of great interest to energy stakeholders. However, it is not uncommon that district heating systems fail to achieve the expected performance due to various faults or inappropriate operations. Night setback is one control strategy, which has been proved to be not a suitable setting for well-insulated modern buildings in terms of both economic factors and energy efficiency. From the literature, shapelets algorithms not only provide interpretable results but also proved to be effective in time series classification. However, they have not been explored to solve the problem in energy domain. In this study, a feature augmentation approach is proposed based on learning time series shapelets and shapelet transformation, aiming to improve the performance of classifiers for night setback classification. To evaluate the effectiveness of the proposed approach, data of 10 anonymous substations in Sweden are used in the case study. The proposed method is applied to six commonly used baseline classifiers: Support Vector Classifier, Multilayer Perceptron Neural Network, Logistic Regression, K-Nearest Neighbor, Decision Trees, and Random Forest. Precision, recall, and f1 score are used as the performance measures. The results of out-of-sample testing show that it is possible to improve the generalization ability of classifiers by applying the proposed approach. In addition, the highest f1 score of out-of-sample testing is achieved by DT classifier whose f1 score is increased from 0.599 to 0.711 for identifying night setback case and from 0.749 to 0.808 for identifying nonnight setback case using the proposed feature augmentation approach.

1. Introduction

In Sweden, the market shares of district heating have increased steadily in the past five decades. District heating is the market leader, supplying approximately 55% of the heat market [1]. Therefore, energy efficiency of district heating systems is crucial.

Different heat usage behaviors are reflected in the corresponding heat load patterns. Analyzing such load patterns plays a crucial role in improving the efficiency of district heating systems management and operation [2]. To be more specific, load pattern analysis can be used by utility companies to adjust control strategies, optimize the operations, personalize the demand management for different target consumer groups, and so forth.

Night setback is a control strategy of energy systems. By applying it, the indoor temperature of a building is lowered at night without reducing the comfort level to save energy and thus cost for the user. Such night setback control is widely integrated with the building management system (BMS) of commercial as well as residential buildings. Although daily load patterns may vary, they show a dip in the evening and a sudden peak in the morning that is due to the need of warming up the cooled-off buildings before occupants’ arrival. The cooccurrence of sudden morning peak demand leads to peaks in the district heating system that can
be problematic for the utility company. Therefore, identification of night setback is of great interest to energy stakeholders.

In this study, a new night setback classification approach based on learning time series shapelets and shapelet transformation is proposed. Shapelets were first proposed by Ye and Keogh [3], which are discriminative time series subsequences that best predict the target. Human interpretable insights to the problem domain can be gained from the result, which is one of the major advantages. Then, a learning shapelets algorithm was proposed by Grabocka et al. [4], in which the optimal shapelets are learnt through the process of optimizing a specified objective function rather than searching for shapelets. Shapelet transform algorithm was proposed by Lines et al. [5], which separates the shapelets extraction process from building a decision tree classifier and allows shapelets to be used by any other classifiers.

On one hand, shapelets algorithms have been successfully applied for time series data analysis in different domains, such as traffic events classification and detection of search and rescue patterns. On the other hand, state-of-the-art methods of night setback identification are based on physical model based approach [6], manual heat load pattern analysis [7], applications of theoretical rules based domain expert knowledge [8], and application of the unsupervised learning approach combined with manual analysis [9].

Neither learning time series shapelets nor shapelet transformation has been used for night setback identification of district heating systems in past studies.

The main contributions of this study are applying learning time series shapelets and shapelet transformation for the first time to the district heating system domain. In addition, from the literature, a transformed dataset by learning time series shapelets and shapelet transformation is used as a new set of features to train the model. Since the new features are encoded after transformation, one potential drawback is by feeding only the transformed features to a model, useful information of original features from the data is lost. Therefore, to retain the useful information of original data, a feature augmentation strategy is proposed by combining both the original features and new shapelet features, which potentially leads to a better performance. Experiment results prove that, by applying the proposed approach, it is possible to improve the generalization ability of the classifiers for night setback classification.

To verify the effectiveness of the proposed approach, real operational data of substations in Sweden are used as a case study. Performance of six commonly used baseline classifiers, Support Vector Classifier (SVC) [10], Multilayer Perceptron Neural Network (MLP) [11], Logistic Regression (LR) [12], K-Nearest Neighbor (KNN) [13], Decision Trees (DT) [14], and Random Forest (RF) [15], are evaluated.

The rest of the paper is organized as follows: relevant past studies of shapelets and heat load pattern analysis are reviewed in Section 2, followed by the overview of methodologies involved and the details of the proposed method in Section 3. In Section 4, the dataset used in the experiment is described. Experiment results are presented and discussed in Section 5. Conclusions of this study are presented in the last section.

2. Literature Review

Recent studies of shapelets extraction, shapelet transformation, and load pattern analysis of district heating substations are reviewed below.

In 2014, Hills et al. [16] proposed a single-scan shapelet algorithm that identifies the best k shapelets. Then, the identified shapelets are used to generate a transformed dataset, which contains k features, and each feature is the distance between the shapelet and a time series. The proposed method was tested against 29 datasets, among which 17 were from the open-access UEA and UCR time series classification repository [17]. Experiments showed that combing shapelet transformation with more complex classifiers leads to a higher accuracy than using the embedded shapelets’ tree. Moreover, three new time series classification domains were introduced for the first time in this study: classifications of Beetle/Fly, Bird/Chicken, and Otoliths.

In 2018, Aldshanhani et al. [18] applied shapelet transform for traffic events classification. In this study, the authors applied a modified shapelets selection process by getting domain experts to choose between existing candidates or creating new shapelets using a visual tool. The M25 London Circular road data of January 2017 is used in the case study. Performance of support vector machines (SVM), KNN, DT, Naïve Bayes (NB), RF, rotation forest (RotF), and collective of transformation-based ensembles (COTE) classifiers are evaluated. The two most important conclusions of this study are as follows: first, the selection of a classifier depends on the dataset and a relatively strong model such as SVM does not always yield the best results. KNN and RF achieved better overall results in the experiment. Second, the experiment results proved that applying shapelet transformation does not always improve the performance of classifiers.

In 2019, Kapadais et al. [19] proposed a framework for the detection of search and rescue patterns based on shapelet classification. The proposed framework consisted of a synthetic data generator to generate the vessel trajectories with different manoeuvre types. Besides, a genetic algorithm based shapelets extraction (GENDIS) algorithm is combined with LRs for the trajectories classification. The proposed method showed a promising result of detecting specific search and rescue (SAR) patterns in vessel trajectories. However, the limitations of this study are as follows: the optimality of the extracted shapelets is not guaranteed by the genetic algorithm (GA-) based shapelet algorithm, and only synthetic data are used for training. Besides, only one classifier is tested in the experiment.

Heat load patterns analysis is not thoroughly studied in the literature due to the lack of high-resolution smart meters installation.

In 2010, Goia et al. [20] applied a functional clustering based approach to group daily load curves of the district heating system. Then, functional linear regression models were applied for short-term peak load forecasting. Hourly
heat consumption data of district heating systems in Turin from years 2001 to 2005 were used in the study. There were four clusters identified. To be more specific, the first two clusters represent daily curves of the winter season that show relatively high heat consumption and the third cluster consists of daily curves mainly in October and April, which represent low heat demand, while the curves in the fourth cluster are from Autumn and Spring with medium heat consumption. However, different load patterns were not identified by the clustering results as shapes of daily curves in different clusters were similar and the main difference between different clusters is the heat consumption level. In addition, the object of the study was short-term peak load forecasting rather than load pattern analysis.

In 2013, Gadd and Werner [7] presented a study regarding the analysis of heat load patterns of district heating substations. Hourly meter readings of 141 buildings in the southwest of Sweden are used in this study. Load patterns with four different control settings, namely, continuous operation control, night setback control, time clock operation control 5 days a week, and time clock operation control 7 days a week, are manually identified according to weekly heat load plots. In 2015, the same authors further presented a fault detection approach based on manual analysis of meter readings [8]. There were three categories of faults identified, unsuitable heat load pattern, low average annual temperature difference, and poor substation control. A set of theoretical rules were applied to identify different faults. For example, all buildings with pronounced night setback control were considered as unsuitable heat load patterns. One challenge of performing such analysis is that knowledge about the activities in the buildings is required.

In 2017, Noussan et al. [2] analyzed load patterns of Turin district heating using real operation data from October 2011 to April 2011. Two types of patterns, hourly and daily patterns, were analyzed. The methods used in this study were energy signature and visualizations of heat load curves. Results showed that the main drivers of heat consumption are outdoor temperature and different operation settings such as night switch-off and night temperature setback. In addition, the analysis result showed that night temperature setbacks result in an energy peak in the morning and it is possible to enhance the conversion efficiency by changing the operation setting.

In 2019, Calikus et al. [9] applied a data-driven approach based on the k-shape clustering algorithm for discovering heat load patterns in district heating. The method identified representative patterns of heat load profiles, control strategies being used in a group of buildings. After clustering, results were examined by domain expert, such a process took huge manual effort. Although the proposed approach showed promising results, the domain knowledge applied and the results obtained in this study were not further being leveraged to automate the process of control strategies identification and reduce future manual effort.

From the literature, the shapelets algorithm has undergone two main improvements since the first time it was proposed. The first improvement was replacing the brute search of shapelets with a learning process, which learns the optimal shapelets that optimize a specified objective function, followed by the second improvement, which eliminates the requirement of binding shapelets extraction process with a decision tree classifier. The second improvement, namely, shapelet transformation, enables the shapelets to be used as independent input features for any other classifiers. From the literature, shapelet transformation has a wider usage in the most recent studies due to its flexibility. However, shapelet algorithms have not been used to address research problems in energy domain. Another important finding is that only shapelet features are used in these studies; one drawback of this approach is that using such transformed features only causes the loss of useful information from the original data.

With respect to districting heating, studies of load pattern analysis are scarce and the study focusing on night setback identification is not found in the literature. The methods used in earlier studies before 2015 are either simple visualizations or applications of theoretical rules, while physical modelling and data-driven approaches are used in recent studies. An unsupervised learning method is used in the most recent study [9], which is the only study that utilizes a machine learning approach.

From the literature, the following research gaps are found: shapelet algorithms are not well explored to solve research problems in energy domain. Moreover, load pattern and night setback analysis are not fully studied. In addition, machine learning approaches have not been widely used to perform analysis.

The motivation of this study is to fill the identified research gaps. To be more specific, learning time series shapelets combined with shapelet transformation methods are explored for the first time in energy domain. A straightforward feature augmentation strategy is proposed to cope with the new shapelet feature as well as retaining useful information in original data. A supervised learning approach is proposed for night setback identification, which is the first attempt in this problem domain.

3. Methodology

Shapelets are one-dimensional “time series” subsequences that best predict the target. They are in some sense maximally representative of a class, which were first proposed by Ye and Keogh [3]. Shapelets provide human interpretable insights to the problem domain, which is one of the major advantages. In general, to extract shapelets from time series data, a set of candidates, definition of distance between a shapelet and the time series, and measurement of the discriminatory ability of the candidates are required.

3.1. Overview of Learning Time Series Shapelets and Shapelet Transformation. In the initially proposed shapelets based classifier algorithm, decision trees were utilized to search for the most discriminative shapelets by applying the data split rule recursively according to the distance between candidate shapelets and each time series in the dataset. However, one
4 Advances in Civil Engineering

The drawback of it is the brute force algorithm used for searching shapelets, which requires exhaustively scanning of the candidates extracted from time series data and leads to time inefficiency when the number of candidates is huge. To overcome the drawback, the learning shapelets algorithm was proposed by Grabocka et al. [4], which is more efficient. In addition, the learning time series shapelets algorithm learns the optimal shapelets that optimize a specified objective function rather than searching for shapelets. Details of the learning time series shapelet algorithm are shown in Algorithm 1.

Shapelet transform algorithm was proposed by Lines et al. [5]. Conventional shapelet algorithms are dependent on decision tree models; such embedded approaches limited the usage of shapelets. Shapelet transform algorithm separates the shapelets extraction process from building a decision tree classifier and allows shapelets to be used by any other classifiers. The transformation algorithm can be summarized as follows:

1. Extract the best \( k \) shapelets \( S \) from training data \( T \).
2. Calculate the normalized Euclidean distance \([21]\) of subsequence between each time series \( T_i \) and \( S_{j=1,..,k} \).
3. A transformed dataset is generated using the calculated \( k \) distances. Each attribute of the transformed dataset represents an individual shapelet and the attribute value is the distance between the original time series and the corresponding extracted shapelet.

Details of the Shapelet transformation algorithm are shown in Algorithm 2.

3.2. Overview of the Proposed Method. In this study, a new feature augmentation approach is proposed based on learning time series shapelets and shapelet transformation. Learning time series shapelets algorithm starts by guessing the initial candidate shapelets. After that, a shapelet differentiable model and a loss function are defined, which enable the shapelets to be learned to minimize the classification loss iteratively in a stochastic gradient descent optimization fashion. Then, the learnt shapelets are used for shapelet transformation.

To avoid overfitting, shapelets extraction is performed using only the training data; then, the extracted shapelets can be applied to both training and testing data for shapelet transformation. As a result, the transformed dataset can be used as independent input features by any classifiers.

The proposed approach takes the advantages of both efficiency and the stochastic gradient descent optimization ability. Meanwhile, the transformed dataset can be further used by other classifiers.

Six commonly used baseline classifiers, Support Vector Regression (SVR), Multilayer Perceptron (MLP), Linear Regression (LR), K-Nearest Neighbor (KNN), Decision Tree (DT), and Random Forest (RF), are used in this study. The performance of each baseline model using original features and after feature augmentation is evaluated.

Experiments are performed in two stages to verify the effectiveness of the proposed approach. To be more specific, 5-fold cross-validation is performed at stage one and out-of-sample testing is performed to further verify the generalization ability of the model on unseen substations data. The overall process of the proposed approach for 5-fold cross-validation is shown in Figure 1, which consists of 4 stages.

3.2.1. Data Preprocessing. First, the entire dataset is normalized and split into training and testing sets. The training dataset is further divided into training and validation sets for 5-fold cross-validation, while the testing data contains only out of samples data that have not been used during either training or validation process. Details of the data and the preprocessing step are presented in Section 4.1.

3.2.2. Learning Shapelets and Shapelets Transformation. After data partition and preprocessing, the training set is fed into the learning time series shapelets model to extract shapelets. Original input contains 24 features, each represents the heat energy usage within one hour of the day. Then, the learnt shapelets are used to perform shapelet transformations for both training and validation sets.

3.2.3. Feature Augmentation. After shapelet learning and transformation, the transformed datasets consist of the distance between each learnt shapelet and the original time series. Then, the transformed dataset is used as a new set of features and concatenated together with the features of the original time series to build the models. By combining both original features and shapelet features, information within the original dataset is retained.

3.2.4. Training and Validation. The augmented input data are used to train and validate the model using 5-fold cross-validation. The performance of six commonly used classifiers, namely, DT, SVC, MLP, RF, LR, KNN, is evaluated.

After the cross-validation, an out-of-sample test is performed in stage two. The process of out-of-sample testing is similar to that of 5-fold cross-validation. To be more specific, the entire training data is used to train a learning time series shapelets model. Then, shapelet transformations are performed for both training and test data using the learnt shapelets. After that, input features are augmented by the transformed dataset, which is used to train and test the six classifiers.

4. Experiments

An anonymized dataset based on measurement data from the primary side of district heating substations in Sweden is used in the case study. The dataset is accessible on request from the utility company Utilifeed [22] using the developed web API and contains measures of hourly resolution and includes the following variables: energy (sum for each hour), volume flow (sum for each hour), supply temperature (instantaneous value at each full hour), and return temperature.
(instantaneous value at each full hour). There are in total ten substations confirmed by domain experts used in this study. Of these, five have been identified by experts as exhibiting night setback characteristics, while five are identified as not having night setback.

An overview of the original data is shown in Table 1. A relatively conservative data split approach is used in this study. To be more specific, six out of ten substations are used for training and validation, while another four unseen substations are used for out-of-sample testing.

For each substation, the original series of one-year length is sliced into daily subseries. Each daily subseries consists of 24 time steps (i.e., 24 hours) for data labelling.

In addition to the heat energy usage of substations, the corresponding hourly weather data are used for data filtering. As the night setback only affects space heating load, the change in load due to night setback is less obvious or nonexistent on warm days. Therefore, the average daily temperature is calculated and a threshold of 10 degrees is applied to filter out daily series whose average temperature exceeds the threshold. Moreover, for certain substations, night setback is applied to workdays only and, therefore, series of weekend days are filtered out as well.

After data filtering, each daily series is labeled in such a way that the daily series that present the characteristics of night setback are labeled as “0” and the remaining daily series are labeled as “1”. Therefore, the filtered and labeled daily series of each substation are randomly shuffled and merged into one big dataset. In addition, the dataset is normalized, so that the mean value of each series is 0 and the standard deviation is 1 for the ease of training. An overview of the preprocessed data is shown in Table 2.

After data preprocessing, there are 996 daily series and 654 series for training/validation and testing following the process of 5-fold cross-validation and out-of-sample testing described in Section 3. One common drawback found in the literature is that only the precision of classifiers is reported. However, precision is not always the most important performance measure, that is, the cost of false alarm and that of a missed-out truth varies from different domains. Therefore, precision, recall, and F1 score [23] are used as reported in 5-fold cross-validation and out-of-sample testing.

Six commonly used classifiers are used in the experiment to evaluate the performance of the feature augmentation approach by learning time series shapelets transformation. The same experiment is performed for both models with original features only and with features augmented.
5. Results and Discussion

The performance of each of the baseline models, SVM, MLP, LR, KNN, DT, and RF, using original features and after feature augmentation is evaluated. Precision, recall, and f1 score are used as performance measures. The averaged 5-fold cross-validation results of each classifier without features augmentation are shown in Table 3.

Results show that SVC outperforms the rest of the models in terms of average f1 score, while DT has the lowest f1 scores among all six models. The 5-fold cross-validation results of each classifier with features augmented are shown in Table 4.

Regarding the performance of the six classifiers with extra shapelet features, f1 scores of MLP are the highest among all models, while f1 score of DT is relatively low. The highest f1 scores of MLP with features augmented are slightly higher than that of SVC with only original features, that is, 0.006 and 0.005 lower for label 0 and label 1, respectively. In addition, f1 scores of the worst and the best classifier differ slightly (~0.05) in both cases. Moreover, for each classifier in both cases, f1 scores of the two labels are almost balanced, with a maximum difference of 0.002. To sum up, for the 5-fold cross-validation, the difference between applying shapelets feature augmentation and using only original features for different classifiers is not significant.

Five-sample daily series are randomly selected from both night setback and nonnight setback cases. The selected five samples represent three types of heat load profiles: a night setback profile, multifamily hot water profile, and random profile. The selected sample daily series of heat load before normalization and after normalization along with the corresponding learnt shapelets are plotted in Figures 2–6.

Figures 2 and 3 show two typical cases of night setback profile, in which, from the figures, there are learnt shapelets

![Figure 1: The workflow of the proposed approach for training and validation.](image-url)
matching the locations where there appears to be a peak in the early morning and a dip in the evening. To be more specific, Figure 2 shows that the learnt shapelets 3, 5, and 6 are located around 6 a.m. of the daily curve to match the corresponding morning peak, while the learnt shapelets 2, 4, and 7 are located at 9 p.m. to match the evening dip. Similarly, there are corresponding learnt shapelets matching the morning peak and evening dip as shown in Figure 3. One important insight is that although the hour of morning peak and evening dip from different daily series vary, such dynamic changes are well captured by the learnt shapelets. As shown in Figure 3, there is a morning peak at 3 a.m. and an evening dip in 7 p.m. which are different from those in Figure 2. As a result, the locations of the learnt shapelets 5 and 6 are at 3 a.m., while shapelet 7 is shown at 7 p.m. to cope with the changes accordingly.

Figures 4 and 5 present two cases from a typical hot water profile of a multifamily residential building featured by presenting a peak in the evening and another one in the evening. From Figure 4, there is a peak at 5 a.m. and another one at 5 p.m. The corresponding learnt shapelet 5 matches the morning peak, while the learnt shapelets 1, 2, and 7 are located around the evening peak. Figure 5 shows another hot water profile sample with both morning and evening peaks showing at different hours of a day. The locations of the corresponding learnt shapelets are changed accordingly to cope with the changes.

Figure 6 shows a sample case of a random profile; in the sense, the corresponding learnt shapelets distribution is difficult to be interpreted in this case. There may exist relatively strong correlations between data used in the 5-fold cross-validation. The main reason is

Table 1: Overview of the original data.

<table>
<thead>
<tr>
<th>Substation</th>
<th>Night setback</th>
<th>Year</th>
<th>Length</th>
<th>Training/validation</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substation 1</td>
<td>n</td>
<td>2017</td>
<td>8760</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Substation 2</td>
<td>n</td>
<td>2017</td>
<td>8760</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Substation 3</td>
<td>n</td>
<td>2017</td>
<td>8760</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Substation 4</td>
<td>n</td>
<td>2017</td>
<td>8760</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Substation 5</td>
<td>n</td>
<td>2017</td>
<td>8760</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Substation 6</td>
<td>y</td>
<td>2017</td>
<td>8760</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Substation 7</td>
<td>y</td>
<td>2017</td>
<td>8760</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Substation 8</td>
<td>y</td>
<td>2017</td>
<td>8760</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Substation 9</td>
<td>y</td>
<td>2016</td>
<td>8784</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Substation 10</td>
<td>y</td>
<td>2016</td>
<td>8784</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Overview of the preprocessed data.

<table>
<thead>
<tr>
<th>Night setback label</th>
<th>Amount of training/validation data</th>
<th>Amount of testing data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (night setback)</td>
<td>498</td>
<td>322</td>
</tr>
<tr>
<td>1 (nonnight setback)</td>
<td>498</td>
<td>332</td>
</tr>
<tr>
<td>Total/percentage</td>
<td>996 (60% of total)</td>
<td>654 (40% of total)</td>
</tr>
</tbody>
</table>

Table 3: Averaged 5-fold cross-validation results of each classifier with original features.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Averaged 5-fold cross-validation results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>precision_label0</td>
</tr>
<tr>
<td>DT</td>
<td>0.921</td>
</tr>
<tr>
<td>KNN</td>
<td>0.934</td>
</tr>
<tr>
<td>LR</td>
<td>0.953</td>
</tr>
<tr>
<td>MLP</td>
<td>0.959</td>
</tr>
<tr>
<td>RF</td>
<td>0.946</td>
</tr>
<tr>
<td>SVC</td>
<td>0.980</td>
</tr>
</tbody>
</table>

Table 4: Averaged 5-fold cross-validation results of each classifier with feature augmentation.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Averaged 5-fold cross-validation results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>precision_label0</td>
</tr>
<tr>
<td>DT</td>
<td>0.923</td>
</tr>
<tr>
<td>KNN</td>
<td>0.938</td>
</tr>
<tr>
<td>LR</td>
<td>0.968</td>
</tr>
<tr>
<td>MLP</td>
<td>0.976</td>
</tr>
<tr>
<td>RF</td>
<td>0.922</td>
</tr>
<tr>
<td>SVC</td>
<td>0.974</td>
</tr>
</tbody>
</table>
that although each series is measured from independent different days, one daily series may correlate with other daily series if they are measured from the same substations and shows a more similar energy usage pattern. In this case, the results of the 5-fold cross-validation reflect the performance of a model that is used to identify the changes in the night setback setting of the same substation overtime.

To remove the potential correlations, unseen substations data are used in out-of-sample testing. The result is of higher importance than the 5-fold cross-validation since out-of-sample testing results reflect the generalization ability of models on unseen data of unseen substations. The testing results of models using original features and those with shapelets features augmented are shown in Tables 5 and 6.

Out-of-sample testing results show that the performance of the classifiers shows an overall degradation comparing with the results of 5-fold cross-validation tests. One reason is that a relatively conservative data partition approach is applied, that is, only 60% of data for training/validation and another 40% of data for testing. Regarding the performance of each classifier, DT with shapelets feature augmentation outperforms the rest by achieving an f1 score of 0.808 for classifying nonnight setback and 0.711 for classifying night setback. Although RF and KNN classifiers using only

Figure 2: Sample case one (night setback profile): morning peak at 6 a.m. and evening dip at 9 p.m.

Figure 3: Sample case two (night setback profile): morning peak at 3 a.m. and evening dip at 7 p.m.
original features show a reasonably good f1 score when
classifying nonnight setback, the performance of identifying
night setback daily series is relatively poor. On the other
hand, the DT classifier with shapelets feature augmentation
shows a more balanced performance in classifying both
night setback and nonnight setback cases. The difference in
f1 score between classifying two labels is less than 0.1. Al-
though the results show that, by augmenting features using
learning time series shapelets and shapelet transformation, it
is possible to boost the performance of an individual clas-
sifier, the improvement of shapelets feature augmentation is
not always guaranteed. To be more specific, the performance
of DT and RF classifiers with features augmented is
improved, while the other four classifiers perform better
using only original features for the out of samples testing,
which is aligned with the conclusion reported in [24].
Therefore, it is advised to test the feature augmentation
approach on various models.

Another finding is that it is important to strike a balance
between training/validation and out-of-sample performance
when selecting a model to ensure the generalization ability to
classify unseen samples. From the experiment results, a
relatively strong model such as SVC and MLP achieves
satisfying results in 5-fold cross-validation testing. However,
the performance of SVC and MLP suffers when the amount
of training is relatively small. As a result, these two models
achieve either a high precision along with a low recall score or vice versa according to the out-of-sample testing results.

In addition to precision, recall, and f1 scores, time measurements of the proposed approach and models without shapelet feature augmentation for cross-validation and out-of-sample testing are reported in Tables 7 and 8.

Results show that the proposed approach is relatively time-consuming due to the extra learning time series shapelets and shapelet transformation processes involved.

Finally, there are some potential improvement areas. For example, the parameters used in the learning time series shapelets, such as the number of shapelets to be learned and length of the shapelet, can be further fine-tuned with evolutionary algorithms (EA) optimization algorithms, such as GA, particle swarm optimization (PSO), and differential evolutionary (DE) [25]. Besides, in addition to applying the proposed feature augmentation approach to an individual

---

**Table 5: Out-of-sample testing of models with original features.**

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>precision_label0</th>
<th>recall_label0</th>
<th>f1_label0</th>
<th>precision_label1</th>
<th>recall_label1</th>
<th>f1_label1</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>0.830</td>
<td>0.469</td>
<td>0.599</td>
<td>0.638</td>
<td>0.907</td>
<td>0.749</td>
</tr>
<tr>
<td>KNN</td>
<td>0.925</td>
<td>0.385</td>
<td>0.544</td>
<td>0.619</td>
<td>0.970</td>
<td>0.756</td>
</tr>
<tr>
<td>LR</td>
<td>0.879</td>
<td>0.317</td>
<td>0.466</td>
<td>0.591</td>
<td>0.958</td>
<td>0.731</td>
</tr>
<tr>
<td>MLP</td>
<td>0.878</td>
<td>0.245</td>
<td>0.383</td>
<td>0.569</td>
<td>0.967</td>
<td>0.717</td>
</tr>
<tr>
<td>RF</td>
<td>0.903</td>
<td>0.404</td>
<td>0.538</td>
<td>0.624</td>
<td>0.958</td>
<td>0.755</td>
</tr>
<tr>
<td>SVC</td>
<td>0.953</td>
<td>0.252</td>
<td>0.398</td>
<td>0.576</td>
<td>0.988</td>
<td>0.728</td>
</tr>
</tbody>
</table>

---

**Table 6: Out-of-sample testing of models with feature augmentation.**

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>precision_label0</th>
<th>recall_label0</th>
<th>f1_label0</th>
<th>precision_label1</th>
<th>recall_label1</th>
<th>f1_label1</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>0.925</td>
<td>0.578</td>
<td>0.711</td>
<td>0.700</td>
<td>0.955</td>
<td>0.808</td>
</tr>
<tr>
<td>KNN</td>
<td>0.919</td>
<td>0.354</td>
<td>0.511</td>
<td>0.608</td>
<td>0.970</td>
<td>0.747</td>
</tr>
<tr>
<td>LR</td>
<td>0.899</td>
<td>0.220</td>
<td>0.354</td>
<td>0.563</td>
<td>0.976</td>
<td>0.714</td>
</tr>
<tr>
<td>MLP</td>
<td>0.857</td>
<td>0.224</td>
<td>0.355</td>
<td>0.561</td>
<td>0.964</td>
<td>0.710</td>
</tr>
<tr>
<td>RF</td>
<td>0.967</td>
<td>0.457</td>
<td>0.620</td>
<td>0.651</td>
<td>0.985</td>
<td>0.784</td>
</tr>
<tr>
<td>SVC</td>
<td>0.947</td>
<td>0.220</td>
<td>0.358</td>
<td>0.566</td>
<td>0.988</td>
<td>0.720</td>
</tr>
</tbody>
</table>

---

**Table 7: Time measurements of 5-fold cross-validation.**

<table>
<thead>
<tr>
<th>5-fold cross-validation time (sec)</th>
<th>Classifier</th>
<th>With feature aug.</th>
<th>Without feature aug.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVC</td>
<td>137.29</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>KNN</td>
<td>143.89</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>DT</td>
<td>148.28</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>156.70</td>
<td>1.98</td>
<td></td>
</tr>
<tr>
<td>MLP</td>
<td>165.25</td>
<td>7.38</td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>166.66</td>
<td>0.68</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 6: Sample case five (nonnight setback, random profile).**
classifier, an ensemble approach [26] or a relatively strong classifier such as deep neural networks can also be explored. Although the proposed approach shows a promising result in this study, further verification using a relatively large number of substations datasets is considered in our future work.

6. Conclusions

In this study, a new feature augmentation approach is proposed based on learning time series shapelets and shapelet transformation algorithms. The proposed approach is applied to six baseline classifiers to identify night setback of district heating systems using the corresponding daily series data. Data from ten anonymous substations of Sweden are used in the case study. Two sets of experiments are designed to evaluate the effectiveness of the proposed approach. The 5-fold cross-validation results show that MLP with shapelets feature augmented and SVC classifier using only original features outperform the rest of the models in terms of f1 score. To further test the generalization ability of each model on unseen substation data, out-of-sample testing is performed. The highest f1 score of out-of-sample testing is achieved by DT classifier whose f1 score is increased from 0.599 to 0.711 to identify night setback case and from 0.749 to 0.808 to identify nonnight setback case using the proposed feature augmentation approach in out-of-sample testing. Results also show that, by applying the proposed feature augmentation approach, the performance of an individual can be potentially improved. In addition, the proposed approach can be used to improve the model performance on the poorly classified label and leads to a relatively balanced f1 score of different labels. However, the improvement is not always guaranteed by applying the proposed approach. Therefore, a thorough test should be performed when considering a model to be used. Finally, potential improvements and future works, including fine-tuning parameters of learning time series shapelets, exploring different data splitting strategies, and using larger datasets, are discussed.

Data Availability

An anonymized dataset based on measurement data from the primary side of district heating substations in Sweden is used in the case study. The dataset is accessible on request from the utility company Utilifeed using the web api (http://www.utilifeed.com).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


