Automatic classification of harmonic data using k-means and least square support vector machine

Hüseyin ERİŞTİ[1,*], Vedat TÜMEN[2], Özal YILDIRIM[2], Belkıs ERİŞTİ[1], Yakup DEMİR[3]

[1]Department of Electrical and Electronics Engineering, Faculty of Engineering, Mersin University, Mersin, Turkey
[2]Tunceli Vocational School, Tunceli University, Tunceli, Turkey
[3]Department of Electrical and Electronics Engineering, Faculty of Engineering, Fırat University, Elazığ, Turkey

Abstract: In this paper, an effective classification approach to classify harmonic data has been proposed. In the proposed classifier approach, harmonic data obtained through a 3-phase system have been classified by using k-means and least square support vector machine (LS-SVM) models. In order to obtain class details regarding harmonic data, a k-means clustering algorithm has been applied to these data first. The training of the LS-SVM model has been realized with the class details obtained through the k-means algorithm. To increase the efficiency of the LS-SVM model, the regularization and kernel parameters of this model have been determined with a grid search method and the training phase has been realized. Backpropagation neural network and J48 decision tree classifiers have been applied to the same data and results have been obtained for the purpose of comparing the performance of the LS-SVM model. The real data obtained from the output of distribution system have been used to assess the performance of the proposed classifier system. The obtained results and comparisons suggest that the proposed classifier system approach is quite efficient at classifying harmonic data.

Key words: Harmonic, measurement, data mining, LS-SVM, k-means

1. Introduction

Harmonics create many negative effects on power system elements and the electrical devices that are connected to the system. In order to eliminate these negative effects, harmonics have to be below the limit values specified in harmonic standards. For this reason, harmonics always have to be monitored in power systems and their harmonic elimination has to be done when required. Harmonics is perhaps the most commonly used word in the electric sector, although their effects and reduction methods are rarely known. Since power quality analyzers are commonly used and can retain data permanently, the situations that occur in power systems can be controlled real-time [1]. Electric distribution network service providers install harmonic monitoring equipment that calculates three-phase harmonic voltage and current details in order to determine and reduce ever-increasing harmonic breakdown problems [2]. For this reason, there has been an increase in the usage of harmonic monitoring systems in power systems in recent years. There are various studies that have been carried out on power quality.

In [3], a new method for power signal harmonic analysis based on frequency and phasor estimating algorithm, a finite-impulse-response comb filter, and a correction factor were proposed. In [4], an ESPRIT-

*Correspondence: huseyineristi@gmail.com

based approach was used for the estimation of interharmonic frequencies of power system current and voltage signals. In [5,6], various power quality monitoring and power quality analyses were proposed. In [7], a new hybrid LS-Adaline algorithm for estimation of harmonic parameters in a power system were presented. In [8], a fast and accurate approach for real-time estimation of moderate time-varying harmonics of voltage and current signals was presented. The proposed method was based on estimation of signal parameters via an ESPRIT-assisted adaptive wavelet neural network. In [9], a spectral decomposition based approach for calculating interharmonic frequencies in power systems by using Kalman filtering was given.

Various data mining techniques are used on harmonic data so that the volume of the harmonic data obtained from monitoring systems is considerably large and hidden patterns in these data can be discovered. When the studies conducted are examined, in [10], clustering of similar groups in a harmonic database was realized by using the ACPRO data mining tool. In [11], a study to determine the number of optimum clusters in clustering harmonic data by using the minimum message length technique was done. In [12], the design of a C4.5 classifier-based online system for the harmonic circumstance monitoring of distribution cables was given. In [13], a system for the monitoring of multiple harmonic sources in power systems by using artificial neural networks was presented.

Each cluster obtained by using clustering techniques for harmonic monitoring data represents a special situation. Various analyses and solutions in parallel with the situations represented by these clusters can be performed by operation engineers. Thus, the clusters representing power quality problems can be used to determine future power quality problems [11].

In this study, a k-means and least squared support vector machine (LS-SVM)-based approach in clustering harmonic data is proposed. This approach, in the first place, involves clustering harmonic data k times with the k-means algorithm. The clusters including harmonic data are also kept as class details for these data. In accordance with the obtained class details, the LS-SVM classifier is trained and harmonics are classified. The harmonic data obtained from the outlet of the distribution system of the Faculty of Engineering at Tunceli University are used to assess the performance of the proposed classifier system.

Harmonic classification has been done with the weekday values of the measured harmonic data. The obtained results show that the proposed system is an efficient approach to use for harmonic classification.

The contributions presented in this paper can generally be summed up as follows:

- An effective harmonic classification by hybrid usage of supervised and unsupervised learning approaches has been presented to classify harmonic data, an important research area in power quality.

- The harmonic data used in the study have been obtained from a real power system over a 3-month continual measurement period. The proposed classification method was applied on these real harmonic data. When the classification results are assessed, it becomes obvious that the proposed approach is an effective approach that can be used on real power systems.

- The proposed classification system has been designed in a way that can be applied separately on a 3-phase system. Thus, an effective approach has been presented to determine harmonic differences between phases in the system with classification.

2. Preliminaries

2.1. Data mining

Data mining is one of the fast developing fields in the computer industry. Although it first appeared as a field under statistics and computer sciences, it has shown a fast development and emerged as a separate field in its

own right. Data mining is the process by which patterns out of huge amounts of data that help us predict the future are sought by means of computer programs and information is obtained. In other words, data mining is the process of finding patterns and relations from data for the purpose of consistent predictions and by using various analytical tools. One of the most important advantages of data mining is that it allows the usage of the most suitable techniques in solving a problem with a wide variety of methodologies and techniques [14].

Data mining models can be studied under three main groups depending on the functions they perform:

- Classification and regression,

- Clustering,

- Association rules and sequential patterns.

Classification and regression models are estimator models; clustering and association rules and sequential patterns are descriptive models. Classifying is a process of finding distinguishing or expository models (or functions) for data classes. This model or function is used to estimate the classes of the objects whose class information is unknown. A derived model is based on the analysis of the training data cluster. The derived model can be represented in various ways such as classification criteria (IF-THEN), decision trees, mathematical formulae, or neural networks. Unlike classification or estimation, clustering analyzes data objects without referring to a known class information. Generally, class information does not exist in training data because this information is unknown. Clustering can be used to produce such information. Objects are clustered in compliance with the principle of maximizing or minimizing interclass similarities [15].

2.2. k-Means clustering method

Clustering is separating elements into a certain number of sets so that similar data are in the same sets. The k-means algorithm is the most commonly used clustering algorithm in data mining. It was first introduced by MacQueen in 1967 [16]. The k-means algorithm is used to determine the cluster centers $(c_1, c_2, \ldots, c_k)$ to which each data point (x_i) is the nearest with respect to the quadratic distance sum. This distance is calculated as follows:

$$Distortion\, D = \sum_{i=1}^{n} \left[\min d\left(x_i, c_k\right)\right]^2 \quad k = 1,\, 2,\, ..,\, K \tag{1}$$

Here d is distance function and the Euclidean distance function is generally used for this calculation. The steps of the traditional k-means algorithm are as follows [17]:

Step1: The value of K, which represents the number of clusters, is determined and the center of each cluster is assigned randomly $(c_1^{(0)}, c_2^{(0)}, \ldots, c_K^{(0)})$. Each cluster center is an m-dimensional vector in the form of $c_i^0 = \left\{ c_{i1}^{(0)}, c_{i2}^{(0)}, \ldots, c_{im}^{(0)} \right\}$

Step2: The distance $d_{ki}^{(l-1)}$ between the ith data set and kth cluster center is calculated with the Euclidean distance equation.

$$d_{ki}^{(t-1)} = \left\| x_i - c_k^{(t-1)} \right\| = \sqrt{\sum_{j=1}^{m} \left(X_{ij} - c_{kj}^{t-1} \right)^2} \tag{2}$$

Step3: Each x_i data point is assigned to the nearest cluster center.

Step4: The cluster centers are updated in accordance with the averages of all the x_i data assigned to the clusters.

Step5: The value of D is calculated with Eq. (1) and if the value of D is closer, the cluster centers are determined and the process ends. If not, iteration is repeated one more time and all the steps are repeated starting from step 2.

Although the k-means algorithm has advantages, it has some shortcomings as well. Its major shortcomings can be listed as follows [18]:

- Selection of initial cluster centers has a considerable effect on the performance of the algorithm.

- The k-means algorithm requires an external constant k parameter.

- The k-means algorithm can only be applied on digital data. It cannot be used for clustering categorical data.

2.3. Least square support vector machines

The LS-SVM was first proposed by Suykens et al. [19]. To speed up SVM training, they took the quadratic function as the empirical risk function and substituted equality constraints for the inequality ones. Instead of solving a quadratic programming problem as in the SVM, the LS-SVM can obtain the solutions of a set of linear equations. The formulation of the LS-SVM are introduced as follows.

Let $\{x_i, y_i\}_{i=1}^N$ be a training set having N number of items and $x_i \in R^n$ as the input data and $y_i \in R$ as the output data. An optimization problem for predicting functions in the LS-SVM can be formulated as follows:

$$J(w, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2, \tag{3}$$

subject to equality constraints

$$y_i = w^T \varphi(x_i) + b + e_i, \quad i = 1, \ldots, N \tag{4}$$

$\varphi : R^n \to R^{n_h}$ represents a function that maps the input data into a feature space with a higher dimension, w the weight vector in the initial weight area, e_i error parameters, and b the bias term.

We introduce the Lagrangian as follows below.

$$L(w, b, e; \alpha) = J(w, e) - \sum_{i=1}^N \alpha_i \left\{ w^T \varphi(i) + b + e_i - y_i \right\} \tag{5}$$

Here the $\alpha = (\alpha_1; \alpha_2; \ldots; \alpha_N)$ Lagrangian is the factor vector. The conditions for optimality are as follows.

$$\frac{\partial L_1}{\partial w} = 0 \to w = \sum_{i=1}^N \alpha_i \varphi(x_i),$$

$$\frac{\partial L_1}{\partial b} = 0, \to \sum_{i=1}^N \alpha_i = 0,$$

$$\frac{\partial L_1}{\partial e_i} = 0, \to e_i = \frac{1}{C} \alpha_i, i = 1, \ldots, N$$

$$\frac{\partial L_1}{\partial \alpha_i} = 0, \rightarrow y_i = w^T \varphi(x_i) + b + e_i, \ i = 1, \ldots, N. \tag{6}$$

After parameters e_i and w are eliminated, the linear equality series regarding the solution is given as follows.

$$\begin{bmatrix} 0 & 1^{\rightarrow T} \\ 1^{\rightarrow T} & \Omega + \gamma^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{7}$$

$y = [y_1; \ldots; y_N]$, $1^{\rightarrow} = [1; \ldots; 1]$, $\alpha = [\alpha_1; \ldots; \alpha_N]$ and Mercer condition

$$\begin{aligned} \Omega_{kl} &= \varphi(x_i)^T \varphi(x_l), \ i, l = 1, \ldots, N \\ &= K(x_i, x_l) \end{aligned} \tag{8}$$

In parallel with the results, the LS-SVM function prediction happens as follows.

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{9}$$

Here α and b are solved with Eq. (7). Various options in the form of core function are available. These are linear, radial basis function (RBF), polynomial, sigmoid, spline, and bspline. RBF and polynomial are the commonly used core functions.

3. Harmonic monitoring system and harmonic data

Monitoring harmonics in power distribution systems is very important in terms of obtaining harmonic data according to harmonic standards and determining harmonic enhancement process. Current harmonic data have been obtained in 10-min intervals specified by the IEC 61000-4-30 standard. The standard regarded as best practice for power-quality (PQ) measurement recommends 10-min aggregation intervals for routine PQ survey. Each 10-min data point represents the aggregate of the ten-cycle root mean square (rms) magnitudes over the 10-min period [20].

In this study, a system for real-time monitoring of the changes in current harmonics in power systems and doing analyses of these data has been designed. The structure of the harmonic monitoring system is given in Figure 1. In this system, current details obtained at the measuring point are transferred to the monitoring device. The harmonic data are obtained real-time by using various algorithms on these signals in the monitoring device. Harmonic data are transferred to a computer by using data transfer technologies that the monitoring device has. Harmonic data are analyzed with monitoring software created on a computer.

Measuring Point   Monitoring Device   PC Software

Figure 1. Harmonic monitoring system.

In order to reduce the harmful effects of harmonics, harmonic values have to be measured correctly, real-time, and continuously. The measurement and analyses of these components are very important in terms of removing the problems in power systems created by harmonics. In the systems being dealt with, at what point the measurement has to be done and how it should be analyzed are matters that require attention.

In this study, the electric distribution panel of the Faculty of Engineering at Tunceli University was used as the center of measurement. Figure 2 gives the real-time harmonic monitoring system installed on the distribution panel of the Faculty of Engineering at Tunceli University. The harmonic data used in this study are taken from the 3 phases on the distribution panel. There are such electrical loads as classrooms, laboratories, analysis lab, heaters, air conditioners, projectors, LCD screens, high-power computers, fluorescent lamps, and security camera systems that are always active at the Faculty of Engineering.
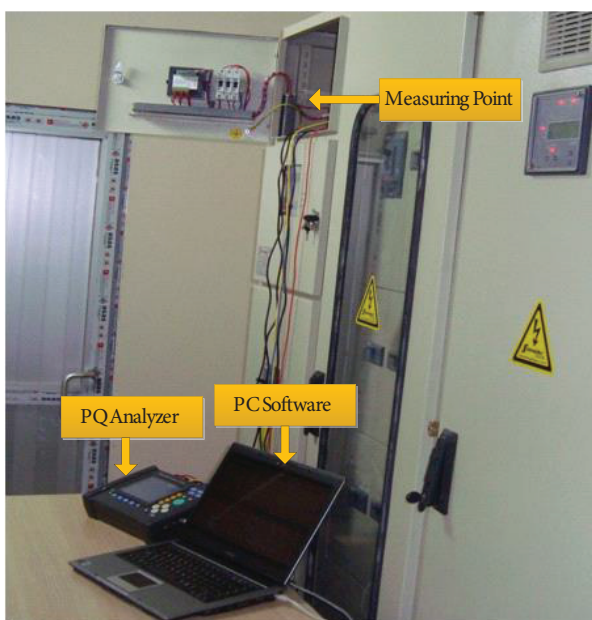

Figure 2. The monitoring system installed on the measurement area.

Current harmonic components on the harmonic monitoring system have been obtained by means of a power quality analyzer. In this study, the 3rd, 5th, 7th, 9th, 11th, and 13th current harmonic components have been used for the proposed monitoring system. The power quality analyzer was connected to the 3 phases on the distribution panel and harmonic data were obtained at 10-min intervals for 12 weeks.

Harmonic data have been divided into two groups, weekdays and weekends, Because electric usage on weekdays and weekends is quite different due to the simple fact that weekends are not office days. Therefore, harmonic classification of the harmonic data pertaining to weekdays, when electrical devices are used extensively, has been conducted in this study. Weekend or overall harmonic classification can be done with the proposed harmonic classification system. The total amount of harmonic data for a single phase is given in Table 1. Since the system measured is a 3-phase system, the number of data to be studied in 12 weeks is 155,520. The distribution of the harmonic degrees for a certain weekday is given separately in Figure 3.

The averages of 12-week harmonic degrees are given in Figure 4. The harmonic degrees given in Figure 4 are seen to be considerably higher than the limit values specified in IEC 61000-4-7/CLASS B.
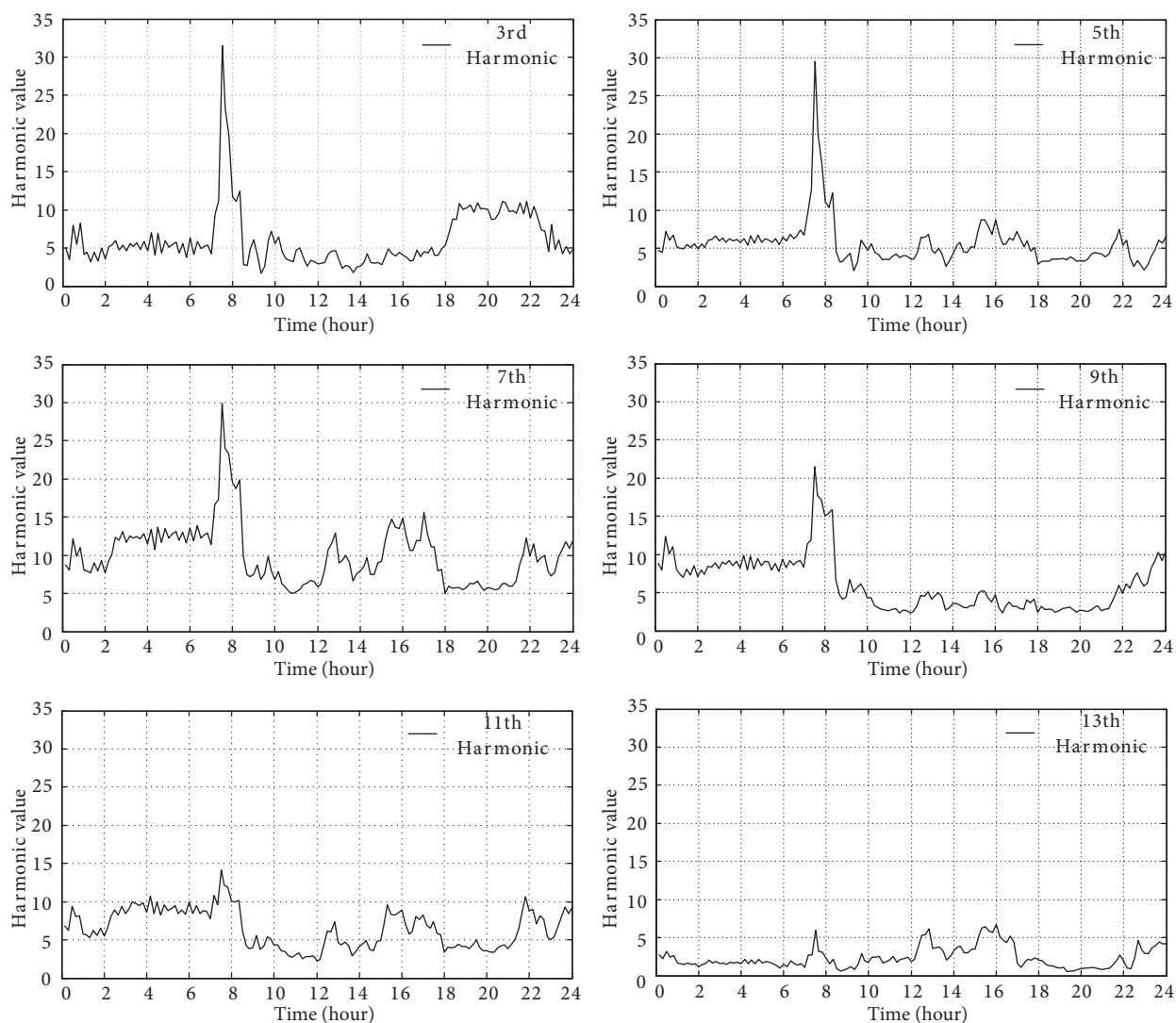
Figure 3. Temporal distribution of harmonic degrees for a randomly selected weekday.

Table 1. The number of data pertaining to harmonic degrees obtained for a single phase.

Harmonic degree	1 hour	1 day	1 week	12 weeks
3rd harmonic	6	144	720	8640
5th harmonic	6	144	720	8640
7th harmonic	6	144	720	8640
9th harmonic	6	144	720	8640
11th harmonic	6	144	720	8640
13th harmonic	6	144	720	8640
Total	36	864	4 320	51,840

4. The proposed classifier approach

In this study, a k-means and LS-SVM-based smart monitoring system that automatically measures the changes in current harmonics in distribution systems has been proposed. The harmonic components of 3-phase current

signals obtained from the output of a power distribution system have been applied to the input of the proposed smart monitoring system and class details pertaining to 3-phase current harmonics have been obtained at the outlet of the monitoring system. The current diagram of the proposed classifier algorithm is given in Figure 5.
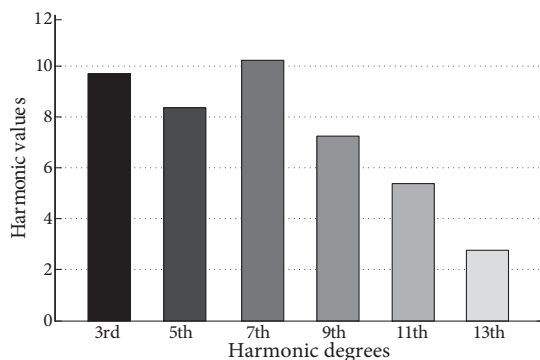

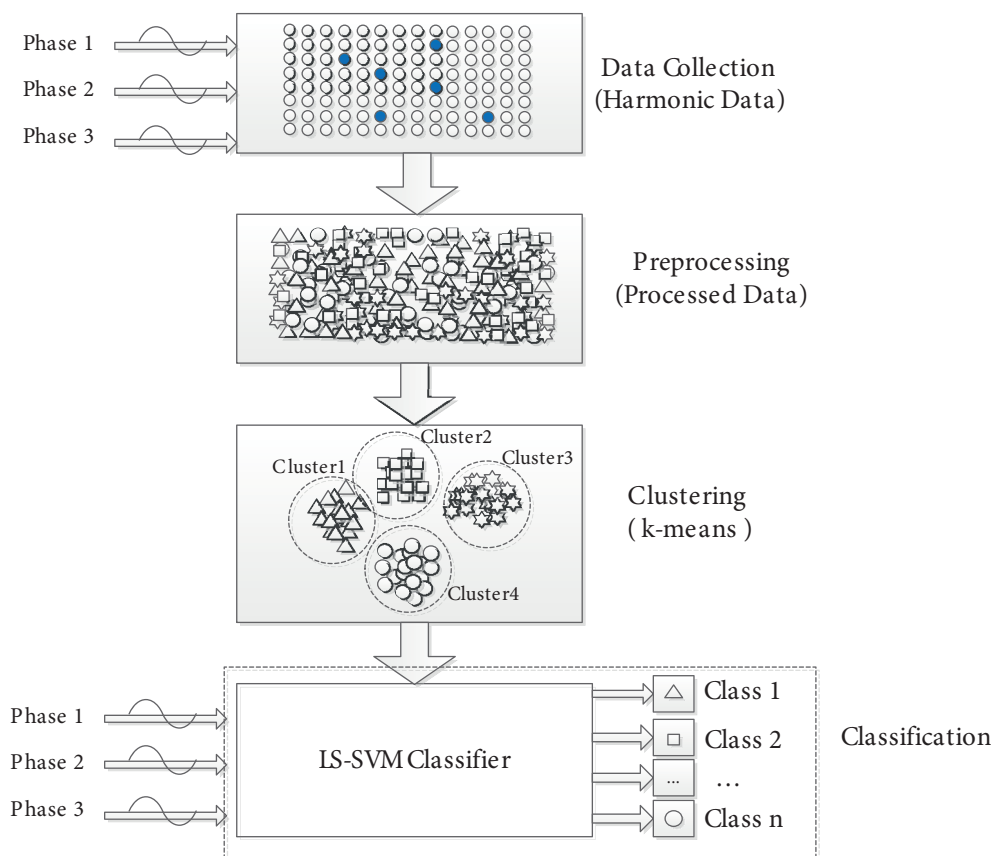**Figure 4.** The averages of weekdays harmonic degrees.

Figure 5. The structure of the proposed harmonic classifier system.

For harmonic data to be classified, they have to have various class details. When the obtained harmonic data are examined, it is clearly seen that they do not have any class details. Therefore, the proposed classifier operates in two important steps: supervised and unsupervised learning.

4.1. Clustering harmonics with *k*-means

In the proposed harmonic classifier system, current harmonics from a 3-phase distribution system are measured first and collected in a database. The collected harmonic data are preprocessed and transferred into a format that clustering and classifier algorithms can process. They are then separated into a predetermined number of clusters with *k*-means. The clusters including the clustered harmonic data are used as class details at the training stage of the classifier.

Figure 6 shows the distribution of harmonic degrees in the 6 clusters of Phase 1 obtained with *k*-means. It is seen that the harmonic degrees in cluster C1 are low and approximate. The harmonic degrees in clusters C2 and C3 are of intermediate levels and those in the clusters in C4, C5, and C6 are the highest for this phase. The aim of clustering is gathering approximate degrees in the same clusters and finding out in how many different ways the harmonics in a system can occur. When looked at with this aim in consideration, harmonic degrees are seen to be falling in different clusters in a logical pattern. We can similarly deduce from Figure 6 that especially the 3rd, 5th, and 7th harmonic degrees in clusters with high values are the highest and have greater effects on the system.
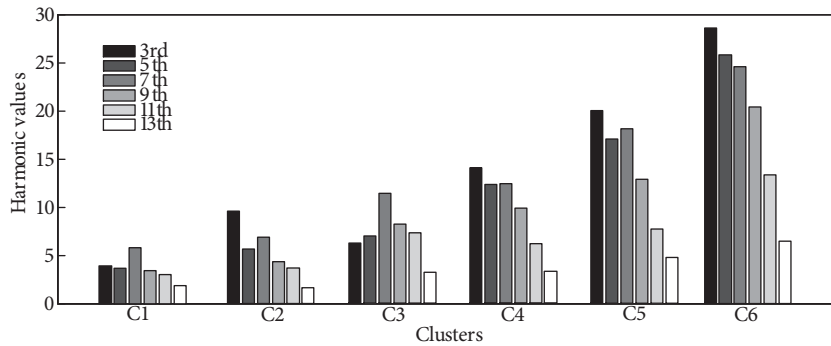

Figure 6. Distribution of harmonic degrees in the 6 clusters of Phase 1.

Figure 7 shows the distribution of harmonic degrees in the 6 clusters of Phase 2 obtained with *k*-means. It is seen that the harmonic degrees in C1, C2, and C3 are low and those in C5 and C6 are the highest for this phase. When the clusters of Phase 2 are examined, it is seen that the harmonic degrees in this phase are lower compared to those in other phases.
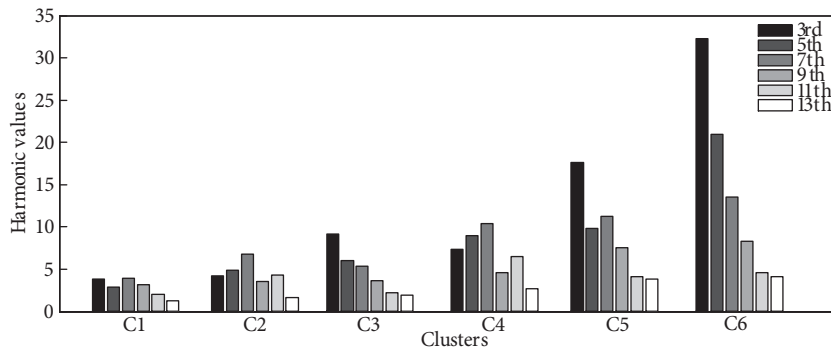

Figure 7. Distribution of harmonic degrees in the 6 clusters of Phase 2.

In Figure 8, the distribution of harmonic degrees for the 6 clusters obtained from the harmonic data with *k*-means is given. A closer look at the figure reveals that the distribution of the harmonic degrees for Phase

3 has a similar structure to the clusters of Phase 1 and has very high values compared to those of Phase 2. The fact that the loads in the phases on the system have different intensities and characteristics may lead to differences in the harmonic values that occur on these phases.
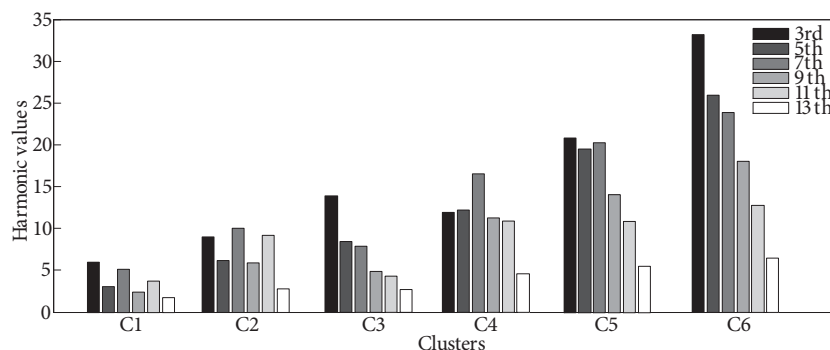

Figure 8. Distribution of harmonic degrees in the 6 clusters of Phase 3.

Table 2 shows the cluster centers that are obtained for the 6 clusters with k-means, which is run harmonic values. When the table is examined, it is seen that the data number in the clusters with low harmonic values is higher than in those with higher levels of harmonics.

Table 2. The cluster centers of phases and the number of data pertaining to these clusters.

Phase 1							
Cluster	Harmonic degree						Data size
	3rd	5th	7th	9th	11th	13th	
C1	3.89	3.61	5.76	3.42	2.97	1.82	2679
C2	9.55	5.58	6.84	4.27	3.67	1.62	1492
C3	6.19	6.91	11.37	8.13	7.24	3.12	1740
C4	13.97	12.21	12.35	9.79	6.09	3.24	1469
C5	19.99	17.01	18.11	12.82	7.63	4.68	939
C6	28.59	25.80	24.55	20.38	13.32	6.39	321
Phase 2							
C1	3.79	2.85	3.80	3.09	1.92	1.20	4345
C2	4.18	4.81	6.66	3.43	4.19	1.57	1983
C3	9.06	5.98	5.23	3.53	2.13	1.81	1229
C4	7.30	8.92	10.34	4.51	6.39	2.63	516
C5	17.50	9.70	11.19	7.45	3.99	3.79	466
C6	32.22	20.87	13.47	8.23	4.54	4.00	101
Phase 3							
C1	5.95	3.02	5.10	2.44	3.73	1.70	2806
C2	8.94	6.17	10.00	5.85	9.20	2.73	1600
C3	13.91	8.37	7.87	4.86	4.25	2.67	1129
C4	11.85	12.15	16.52	11.22	10.85	4.53	1106
C5	20.79	19.52	20.26	14.08	10.82	5.53	815
C6	33.20	25.91	23.85	18.02	12.73	6.41	1184

4.2. The structure of LS-SVM

Some parameters have to be arranged in order to achieve high performance with the LS-SVM. These parameters are the rearrangement parameter (γ) and kernel parameter (σd), which corresponds to kernel type. Adopting

the right method for arranging and optimizing LS-SVM parameters is very important in terms of achieving a good performance and avoiding too much measurement time. Selecting these parameters mostly depends on the type of the work performed. In this study, searching for grids is done on parameter space. Miss-class values from L-fold cross-validation for each grid point are determined and the minimum miss-class and mean squared error (mse) values are measured. L-fold cross-validation has been used in this study because it is a very reliable method for estimating a generalization error [http://www.cs.cmu.edu/∼schneide/tut5/node42.html]. In L-fold cross-validation, training data are randomly divided into L times equal clusters. The LS-SVM classifier model is trained by using these subclusters (L − 1) and verification is done on the rest. This is repeated L times for each of the L subclusters, which are consecutively used for the verification subcluster. The verification error rate that is found as a result of L times trials provides an idea about generalization error. The miss-class equation used for this is as follows:

$$missclass = \frac{1}{N}\sum_{i=1}^{N}|yt_i \neq yp_i|. \tag{10}$$

In this equation, N represents the number of data patterns in the data set, yt the test value, and yp the estimated value.

5. Results and discussion

The k-means clustering algorithms have been used to determine the classes that belong to harmonic data and identify the class centers. Based on the classes determined, the LS-SVM model classifier is applied to estimate harmonic classes. A grid search algorithm is applied to the parameter space for selecting the optimal parameters of the LS-SVM model. A total number of 8640 data, 50% of which are test data and 50% training data, from 12-week harmonic data have been studied for each phase. These data have been clustered with the k-means algorithm and their class details have been determined, and then LS-SVM training has been performed with the randomly selected 4320 data. The k-means and LS-SVM models have been applied separately for all phases. The steps taken for the proposed classifier approach are generally as follows:

- Collecting harmonic data from phases and creating a separate data set for each phase,

- Preprocessing the collected data sets and transforming the data into the presentation formats of the algorithms to be used,

- Separating harmonic data into a specified number of clusters with the k-means algorithm,

- Determining the optimization of the regularization parameter (γ) and kernel parameter (σd) of the LS-SVM model with the grid search method,

- Running the LS-SVM model on test data according to the most suitable value parameters and obtaining the results of classification.

In order to assess the performance of the LS-SVM model, RBF kernel and polynomial kernel functions have been applied on the test data. Similar estimations have been done for both of the kernel functions. Classification and clustering stages have been performed with programs created in MATLAB, and BPNN and J48 classifiers have been run on the same data in order to assess the estimation performance of the LS-SVM model. The results obtained from BPNN, J48, and LS-SVM classifiers have been compared. In this paper, classification of harmonic data with a high accuracy is regarded as the criteria of classifier performance.

Table 3 shows BPNN architecture and training parameters that have been used for performance comparison. BPNN is a commonly used artificial neural network technique used for classification and estimation. One of the most important advantages of BPNN is that it has a very strong nonlinear mapping ability and a flexible network structure. Therefore, it is the most commonly used artificial neural network today. Although the training stage of BPNN classifiers is slow, it stands out among the other artificial neural networks in terms of its classification accuracy. The J48 classifier has also been used to assess the performance of the proposed classifier model. J48, which is based on the C4.5 algorithm, is an algorithm that was first invented in 1993 by Quinlan [21]. It creates a classification decision tree by separating the data into subsets iteratively. The algorithm can be used for the qualities that have continual values, it can prune, and it can be used to produce decisions. The parameters belonging to the J48 algorithm and used in the study have generally been determined as follows: minimum number of instances in a leaf is 2, use of unpruned trees is false, confidence factor used in postpruning is 0.25, use of binary splits is false, and subtree raising operation in postpruning is true.

Table 3. BPNN architecture and training parameters.

Architecture	Number of layers	2
	Hidden layers	30
	Activation functions	Tansig
	Initial weights and biases	Random
Training parameters	Learning rate	0.65
	Moment constant	0.75
	Minimum performance gradient	1e-25

When classification results are examined, it is seen that the LS-SVM model shows a more efficient estimation performance compared to RBF kernel function and other models. Tables 4–6 show classification results of BPNN, J48 classifiers, and the LS-SVM model. It is seen that the performances of classification models on harmonic data for Phase 3 are higher than for other phases.

Table 4. Classification results of harmonic data for Phase 1.

Classifiers	Regularization and kernel parameters		Tested data	Miss-class	Accuracy
	γ	$\sigma, [t; d]$			
RBF kernel (LS-SVM)	2.280	147.81	4320	53	98.77%
Poly-kernel (LS-SVM)	23e-4	[149.23; 3]	4320	619	85.67%
BPNN			4320	135	96.87%
J48 decision tree			4320	192	95.55%

Table 5. Classification results of harmonic data for Phase 2.

Classifiers	Regularization and kernel parameters		Tested Data	Miss-class	Accuracy
	γ	$\sigma, [t; d]$			
RBF kernel (LS-SVM)	96.28	0.704	4 320	81	98.12%
Poly-kernel (LS-SVM)	2e-3	[71.72; 3]	4 320	533	87.66%
BPNN			4 320	131	96.96%
J48 decision tree			4 320	223	94.83%

Table 6. Classification results of harmonic data for Phase 3.

Classifiers	Regularization and kernel parameters		Tested Data	Miss-class	Accuracy
	γ	σ, $[t; d$]			
RBF kernel (LS-SVM)	187.95	94.31	4 320	38	99.12%
Poly-kernel (LS-SVM)	48e-3	[1.45; 3]	4 320	504	88.33%
BPNN			4 320	91	97.89%
J48 decision tree			4 320	247	94.28%

6. Conclusions

In this study, an LS-SVM-based approach to classify harmonic data that occur in a 3-phase system has been proposed. In the proposed approach, harmonic data have been clustered with the k-means algorithm and class details for each data have been obtained. The training of the LS-SVM classifier has been done according to these class details. In order to increase the performance of the LS-SVM model, the most suitable regularization and kernel parameters have been found with the grid search method. In the performance assessment of the LS-SVM classifier, RBF and polynomial kernel functions have been applied separately on test data. The results obtained from BPNN, J48 classifiers, and the LS-SVM model have also been compared in this study. When the results are compared, it is seen that the LS-SVM model has stronger estimation ability than other models. In light of the results and comparisons, it is seen that the proposed classifier has a considerably high performance of harmonic data estimation.

References

[1] Romero Gordon JM. New trends for power quality monitoring: towards a fast response analysis. In: 19th International Conference on Electricity Distribution; 21–24 May 2007; Vienna, Austria. Liege, Belgium: AIM. p. 1–4.

[2] McGranagahn M. Trends in power quality monitoring. IEEE Power Eng Rev J 2001; 21: 3–9.

[3] Yang J, Yu C, Liu C. A new method for power signal harmonic analysis. IEEE T Power Deliver 2005; 20: 1235–1239.

[4] Gu IYH, Bollen MHJ. Estimating interharmonics by using sliding-window ESPRIT. IEEE T Power Deliver 2008; 23: 13–23.

[5] Özgönenel O, Terzi ÜK, Khan A. A hybrid approach for power quality. Turk J Electr Eng Co 2012; 20: 854–869.

[6] Erişti H, Demir Y. The feature selection based power quality event classification using wavelet transform and logistic model tree. Przegląd Elektrotechniczny 2012: 43–48.

[7] Joorabian M, Mortazavi SS, Khayyami AA. Harmonic estimation in a power system using a novel hybrid least squares-Adaline algorithm. Electr Power Syst Res 2009; 79: 107–116.

[8] Sachin K, Singh SN. Fast harmonic estimation of stationary and time-varying signals using EA-AWNN. IEEE T Instr Meas 2013; 62: 335–343.

[9] Köse N, Salor Ö, Leblebicioğlu K. Interharmonics analysis of power signals with fundamental frequency deviation using Kalman filtering. Electr Power Syst Res 2010; 80: 1145–1153.

[10] Asheibi A, Stirling D, Sutanto D. Analyzing harmonic monitoring data using supervised and unsupervised learning. IEEE T Power Deliver 2009; 24: 293–301.

[11] Asheibi A, Stirling D, Sutanto D. Determination of the optimal number of clusters in harmonic data classification. In: Harmonics and Quality of Power ICHQP; 28 September–1 October 2008; Wollongong, Australia. New York, NY, USA: IEEE. pp. 1–6.

[12] Catterson VM, Bahadoorsingh S, Rudd S, McArthur SDJ, Rowland SM. Identifying harmonic attributes from online partial discharge data. IEEE T Power Deliver 2011; 26: 1811–1819.

[13] Negnevitsky M, Ringrose M. Monitoring multiple harmonic sources in power systems using neural networks. In: 2005 IEEE Power Tech; 27–30 June 2005; St. Petersburg, Russia. New York, NY, USA: IEEE. pp. 1–6.

[14] Kantardzic, M. Data Mining: Concepts, Models, Methods, and Algorithms. Chichester, UK: John Wiley & Sons, 2011.

[15] Han J, Kamber M, Pei J. Data Mining Concepts and Techniques. Waltham, MA, USA: Morgan Kaufmann, 2006.

[16] MacQueen JB. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Symposium on Math, Statistics, and Probability. Berkeley, CA, USA: University of California Press, 1967. pp. 281–297.

[17] Kalyani S, Swarup KS. Particle swarm optimization based k -means clustering approach for security assessment in power systems. Expert Syst Appl 2011; 38: 10839–10846.

[18] Zalik KR. An efficient k-means clustering algorithm. Pattern Recogn Lett 2008; 29: 1385–1391.

[19] Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett 1999; 9: 293–300.

[20] IEC. IEC Standard for Electromagnetic Compatibility (EMC)—Part 4-30: Testing and Measurement Techniques— Power Quality Measurement Methods, IEC 61000-4-30. Geneva, Switzerland: IEC, 2008.

[21] Quinlan JR. C4.5: Programs for Machine Learning. San Mateo, CA, USA: Morgan Kaufmann Publishers, 1993.