The Black Sea Journal of Sciences 7(1), 1-8, 2017

KFBD

Karadeniz Fen Bilimleri Dergisi
The Black Sea Journal of Sciences
ISSN (Online): 2564-7377

Research Article / Araştırma Makalesi

K-Nearest Neighbor Classification of Harmonics Using Akaike Information Criterion

Özgür TOMAK[1*], Onur Özdal MENGİ[1]

[1]Giresun University, Department of Energy Systems Engineering, Giresun, Turkey

[*]**Corresponding Author:** ozgur.tomak@giresun.edu.tr

Received: 10.04.2017
Accepted: 03.05.2017

Abstract

If power quality can be maintained, high performance is possible for electrical devices. Harmonics is one of such problems that can cause performance to drop. Therefore, harmonics must be detected and prevented. This study aimed to be part of detection system designed to maintain power quality. Akaike Information Criterion is used to calculate features for power quality analysis. And for classification k-Nearest Neighbor classification is used. %94.8 accuracy is obtained from training set and %91.7 accuracy is obtained from test set. MATLAB is the program which is used for classification and feature calculation.

Keywords: Akaike Information, Harmonic Detection, k-Nearest Neighbor Classification, Power Quality.

Harmoniklerin Akaike Bilgisi Kriterini Kullanarak K-En Yakın Komşu Sınıflandırması

Özet

Elektrik kalitesi korunabiliyorsa, elektrikli cihazlar için yüksek performans mümkündür. Harmonikler, performans düşmesine neden olan bu tür sorunlardan biridir. Bu nedenle, harmonikler tespit edilmeli ve engellenmelidir. Bu çalışma, güç kalitesini korumak için tasarlanan algılama sisteminin bir parçası olmayı amaçlamıştır. Akaike Bilgi Kriteri, güç kalitesi analizi için özellikleri hesaplamak için kullanılır. Ve sınıflandırma için k-En Yakın Komşu sınıflandırması kullanıldı. Eğitim setinden% 94.8 doğruluk elde edilmiştir ve Test setinden% 91.7 doğruluk elde edilmiştir. MATLAB, sınıflandırma ve özellik hesaplama için kullanılan programdır.

Anahtar Kelimeler: Akaike Bilgi Kriteri, Harmonik Tespiti, k-En Yakın Komşu Sınıflandırması, Güç Kalitesi.

1. Introduction

Good quality power is needed by electrical devices to operate properly. Power quality, refers to "maintaining near sinusoidal waveform of power distribution bus voltages and currents at rated magnitude and frequency", determining the fitness of electric power to consumer devices. Thus Power quality is often used to express voltage and current quality, reliability of service, quality of power supply, etc (Chattopadhyay et al., 2011).

Some of the research on this subject is as follows: Lin and Lee, 2001, present an enhanced parametric algorithm which is suitable for online harmonic analysis and performs well with short data. Qian et al., 2007, proposed an interharmonics estimation method based on the interpolation FFT algorithm. Huang et al., 2005, presented a new kind of windows, called as Rectangular Self-Convolution Windows (RSCW). The m-order RSCW is developed by convolving m rectangular sequences and in the zero points of its amplitude-frequency characteristic, the value of m-1 order derivatives is zero. As a result of this character, the interferences between the harmonics due to spectrum leakage can be reduced furthest by applying RSCW, therefore, the precision of harmonics estimation is boosted.

It is of great importance to accurately find out the harmonics component for the safe and economical operation of the power system. Zhan et al., 2004, proposed a novel robust approach to harmonics and interharmonics analysis, based on Support Vector Machines and solved by Iterative Reweighted Least Squares algorithm to overcome the difficulty of exponential computation complexity. Wen et al., 1994, proposed a new power harmonics measurement. The 4-term Blackman-Harris window is used for windowing the sample data so that the spectral leakage errors are largely decreased, and then, dual interpolations are used to obtain harmonic parameters. Jiang et al., 2010, presented the poly-cosine window and double interpolation FFT algorithm to reduce disturbance by the frequency leakage and improve the accuracy of inter-harmonics analysis. The windowed FFT transform is applied to the sampling signal choosing a suitable window function. Then a further amendment is made to the harmonic analysis results using double interpolation algorithm.

In power system, it is especially important for power quality to identify all modes of harmonics and interharmonics. Ganyun et al., 2005, presented a signal decomposition algorithm based on the Prony method on the basis of interharmonic characteristics combining the exponential transform in time-frequency domain with the description of transient signal characteristic. Wang et al., 2008, provided an improved algorithm for analysis of non-integer harmonics in electric power systems by combining the windowed fast Fourier transform (FFT) algorithm with the improved ANN model. Firstly, the Hanning-windowed FFT algorithm processes the sampled signal. By this time, the number of harmonics and the orders of harmonics are obtained.

2. Material and Method

Improvement of power quality has been worked all over the world for decades. Usage of electronic devices has become into widespread and electronic devices has become more sensitive to voltage disturbances. This is a big problem to be solved.

There are lots of power quality problems such as over voltage, frequency harmonics, blackout, short circuit, noise, etc. But harmonics is the problem to be analyzed and classified in this paper. First introduction on harmonics will be given and we will explain database which is used in this research. Next Akaike Information Criterion which will be used for feature extraction is described. Last part is the classification. k-Nearest Neighbor is used for classification. In our case we have 2 classes. First is the sinus wave without harmonics and second is sinus wave with harmonics.

2.1. Harmonics

Any repetitive distorted waveform can be broken down into pure sine waves whose frequencies are integral multiples of the fundamental frequency. These pure sine waves that make up the nonsinusoidal waveform are the harmonic components (Henderson and Patrick, 1994). A waveform with harmonics generated by MATLAB is given in Figure 1.

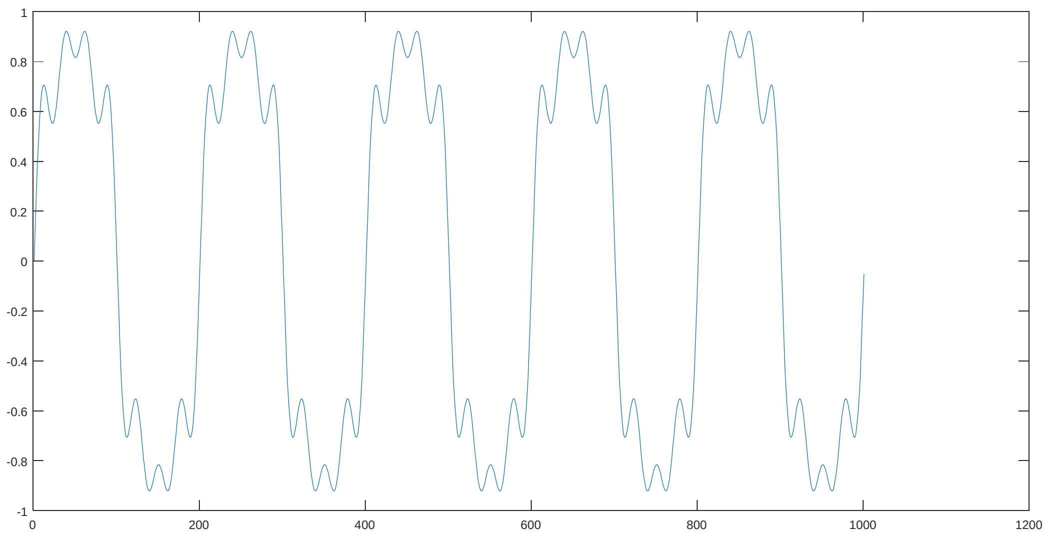

Figure 1. A waveform with harmonics.

There are lots of reasons of why harmonic is produced in the system. Iron core devices is one of the common source of harmonics. Excitation current waveform is not sinusoidal due to the nonlinear hysteresis. Fourier analysis of the current waveform causes a third harmonic component.

As another source of harmonics, Generators produce fifth harmonic voltages because of magnetic flux distortions. Nonlinear loads like inverters, rectifiers, welders, adjustable speed motor drives, voltage controllers, arc furnaces, frequency converters, are also producers of harmonics.

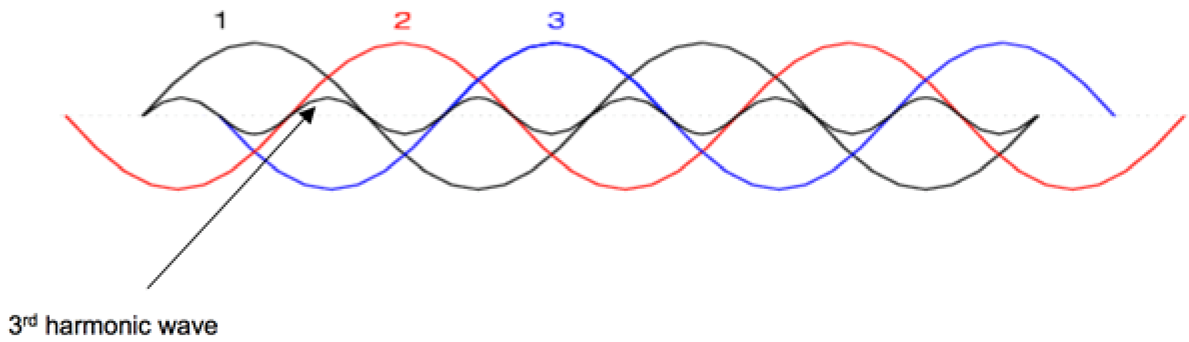

Figure 2. Third order harmonic.

Harmonics are multiples of fundamental wavelength. So, third multiple of the fundamental wavelength is known to be the third order harmonic given in Figure 2. If three phases contain third order harmonics, currents won't fully add to zero. Third harmonic will add with third harmonics within other phases which will cause an oscillating current in neutral wire, Delta connections are used to avoid this dangerous situation.

If we look at some of the effects of the harmonics, first harmonics is known to increase the current in the of power systems. This is due to the third harmonic, which causes an increase in zero sequence current, and increases the current in the neutral conductor. If electric system is to serve non-linear loads, the design requires special consideration.

Losses is experienced by electric motors because of hysteresis and foucault currents. These losses are proportional to the frequency. And harmonics are at higher frequencies, so higher core losses are produced than power frequency would. Increased heating of core may shorten life of the motor. The 5th harmonic causes counter electromotive force (CEMF) in large motors. And this counter electromotive force acts in opposite direction of rotation. Speed of motor reduces due to the CEMF.

2.2. Database

The database used in the method is made of 20 sets. Each set is composed of 1000 signal parts. Waveform with harmonics is randomly distributed. Half of the data is used for training and other half is used for test procedure.

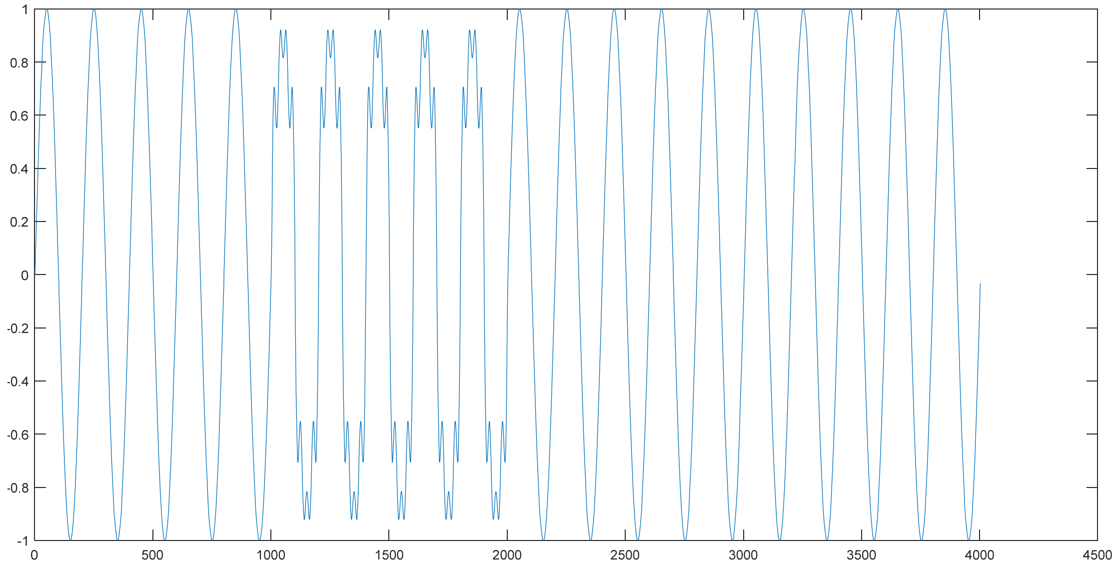

Figure 3. A part of the signal which is used for database.

As can be seen from Figure 3 part of the signal which is used for database, harmonics component is randomly distributed.

3. Akaike Information Criterion

Feature that is used to classify data is Akaike Information Criterion (AIC). A model is chosen from a set of models. The chosen one minimizes Kullback Leibler distance between the truth and model (Burnham and David, 1998). It is defined as

$$AIC = -2(\ln(likelihood)) + 2K \quad (1)$$

K: the number of free parameters of the model

Likelihood: the probability of data of a model

AIC scores are usually shown as ΔAIC scores, or difference of the best model and each model so the best model has a ΔAIC of zero.

4. Classification

Feature extraction for data description is different from feature extraction for classification. Classification based on distance to points in a dataset is a simple but effective. If there is a set of n points, k-nearest neighbor (KNN) method finds the closest points in set to a point or a set of points.

In order to find for classifying each point when predicting, the number of nearest neighbors is to be specified. It is possible to specify a fine classifier or coarse classifier by deciding on the number of neighbors. A fine KNN uses low number of neighbors, a coarse KNN uses high number of neighbors. Using high number of neighbors can be time consuming. In our case the number of neighbors is set to 1 for fast classification. K-nearest neighbor (KNN) classification for sample data is given in Figure 4.

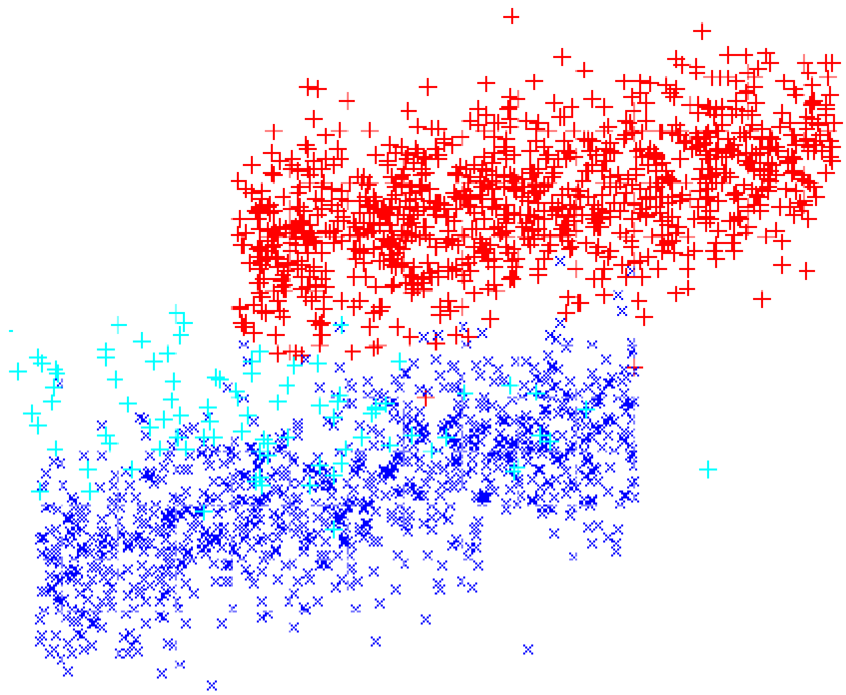

Figure 4. K-nearest neighbor (KNN) classification for sample data.

Different methods can be used to find the distance to points. Distance metric is chosen to be Euclidean for this research.

$$Euclidean\ Distance = \sqrt{\sum(x_i - y_i)^2} \quad (2)$$

In order to specify the distance weighting function, equal, inverse, or squared inverse can be chosen. For equal means no weights, inverse means weight is 1/distance, and squared inverse means weight is 1/distance^2. In our case distance weight is chosen to be equal.

Scaling each coordinate distance is another factor to decide. In our case we chose to standardize data.

In summary for k-Nearest Neighbor Classification, memory usage is medium. Prediction speed is medium. Interpretability is hard. And for Model flexibility, we can say finely detailed distinctions between classes. And model flexibility reduces with the number of neighbors.

Sensitivity refers to correctly detect status which have the condition. Sensitivity measures the proportion of true positives to true positives and false negatives. Specificity relates to correctly detect status without a condition. Specificity measures the proportion of true negatives to true negatives and false positives. Accuracy measures the proportion of true positives and true negatives to true positives and negatives plus false negatives and positives.

True positive (TP) : True condition correctly identified as true condition
False positive (FP) : False Condition incorrectly identified as true condition
True negative (TN) : False Condition correctly identified as false Condition
False negative (FN) : True condition people incorrectly identified as false Condition
These definitions are expressed by the formulas below

$$Sensitivity = TP \,/\, (TP + FN) \quad (3)$$
$$Specificity = TN \,/\, (TN + FP) \quad (4)$$
$$Accuracy = (TP + TN) \,/\, (TP + FP + FN + TN) \quad (5)$$

Table 1. Training and Test Results for K-Nearest Neighbor Classification.

Classification Method	**Training**	**Test**		
	Accuracy	**Accuracy**	**Sensitivity**	**Specificity**
k-Nearest Neighbor Classification	94.8 ±0.46	91.7 ± 2.17	85.8 ±2.26	95.6 ±2.97

5. Conclusion

Waveforms with harmonics are analyzed and classified 5 times. Average values and standard deviations are calculated. Akaike Information Criterion is used for obtaining features and k-Nearest Neighbor is used for classification. k-Nearest Neighbor classification has good predictive accuracy in low dimensions, but might not in high dimensions, and has high memory usage, and is not easy to interpret. Using this system %94.8±0.46 accuracy is obtained from training set and %91.7± 2.17 accuracy is obtained from test set. Sensitivity is found to be 85.8±2.26 and specificity is 95.6 ±2.97.

Our algorithm is fast and it is assumed respond to harmonics fast enough. And accuracy is planned to be increased with a better algorithm. For example number of neighbors can be changed for better results. Also more power quality problems will be added to the research to further increase number of classes.

References

Burnham, K.P., David R. A., (1998) "Introduction." Model Selection and Inference. Springer New York,. 1-31.

Chattopadhyay, S., Mitra, M., Sengupta, S., (2011). Electric Power Quality. Springer Science+Business. ISBN 978-94-007-0634-7.

Ganyun, D., Haozhong L., Lu R., Sun Y., Yifeng, C., Zhan Y., (2005) "Spectrum Estimation of Harmonics and Interharmonics Based on Prony Algorithm [J]." Transactions of China Electrotechnical Society 10: 018.

Henderson, R. D., Patrick J. R., (1994) "Harmonics: the effects on power quality and transformers." IEEE transactions on industry applications 30.3: 528-532.

Huang, C., Ya-qun J., (2005). "Improved window and interpolation algorithm for analysis of power system harmonics." Zhongguo Dianji Gongcheng Xuebao(Proc. Chin. Soc. Electr. Eng.). Vol. 25. No. 15.

Jiang, C., Min L., (2010). "Inter-harmonics analysis based on double interpolation FFT algorithm [J]." Power System Protection and Control 3

Lin, H. C., Lee. C. S. (2001). "Enhanced FFT-based parametric algorithm for simultaneous multiple harmonics analysis." IEE Proceedings-Generation, Transmission and Distribution 148.3: 209-214.

Qian, H., Rongxiang Z., Tong C., (2007). "Interharmonics analysis based on interpolating windowed FFT algorithm." IEEE Transactions on Power Delivery 22.2: 1064-1069.

Qian Y., Wen, P., Zhou. E. (1994). "Power Harmonics Measurement Based on Windows and Interpolated FFT (□) Dual Interpolated FFT Algorithms [J]."Transactions of China Electrotechnical Society 2

Wang, G., Dong-yang X., Wei-ming M., (2008) "Improved algorithm for non-integer harmonics analysis based on FFT algorithm and neural network. "PROCEEDINGS-CHINESE SOCIETY OF ELECTRICAL ENGINEERING 28.4: 102.

Zhan, Y., et al. (2004). "A Robust Support Vector Algorithm for Harmonics Analysis of Electric Power System [J]." Proceedings of the Csee 12: 009.