2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)
8-10 December 2016, Rajshahi-6204, Bangladesh

A New Machine Learning Approach to Select Adaptive IMFs of EMD

Md. Burhan Uddin, Jia Uddin, Razia Sultana and Safat Islam
Dept. of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh
burhan420@gmail.com, jia.uddin@bracu.ac.bd, rsnishi.1992@gmail.com, skyscraper1709@gmail.com

***Abstract*—An adaptive algorithm for selection of Intrinsic Mode Functions (IMF) of Empirical Mode Decomposition (EMD) is a time demand in the field of signal processing. This paper presents a new model of an effective algorithm for the adaptive selection of IMFs for the EMD. Our proposed model suggests the decomposition of an input signal using EMD, and the resultant IMFs are classified into two categories the relevant noise free IMFs and the irrelevant noise dominant IMFs using a trained Support Vector Machine (SVM). The Pearson Correlation Coefficient (PCC) is used for the supervised training of SVM. Noise dominant IMFs are then de-noised using the Savitzky-Golay filter. The signal is reconstructed using both noise free and de-noised IMFs. Our proposed model makes the selection process of IMFs adaptive and it achieves high Signal to Noise Ratio (SNR) while the Percentage of RMS Difference (PRD) and Max Error values are low. Experimental result attained up to 41.79% SNR value, PRD and Max Error value reduced to 0.814% and 0.081%, respectively compared to other models.**

Keywords—intrinsic mode functions; empirical mode decomposition; support vector machine; Pearson's correlation coefficient

I. INTRODUCTION

Signal processing has spread its foot print thorough out most of the technological achievements and advancements. Use of signal processing in home automation to deep space exploration shows its extended reach in our life. Advance robotics, medical research and diagnostics, artificial intelligence all these respective fields has signal processing as backbone since long ago. Signal processing is used acquiring desired information as outcome from a set of raw data [1].

In analysis of non-stationary signals Wavelet transformation has been used over Fourier based transformation in regards of their performances [14]. In recent times EMD (Empirical Mode Decomposition) has proven its might on non-stationary signal processing leaving behind the wavelet transformation with locality and adaptability issues [2,3,6]. EMD decomposes a signal into a finite set of frequency modulated components which are IMFs (Intrinsic Mode Function). The EMD is adaptive and efficient in decomposing a signal. The decomposed IMFs are separated in two sets: noise free IMFs and noise dominant IMFs. For separation of the IMFs different researchers have used different approaches. Phuong and others in their paper have used Naïve Bayes classifier [3], it is a probabilistic and decision based classifier for the separation of relevant and non-relevant IMFs. On contrary to that, in paper [5] and [6] have proposed an energy based thresholding and classification method. We have followed the thresholding technique proposed by [6] for our SVM training. As the comparison results [4] show it as being better.

There are several attractive features in a SVM. Due to its excellent empirical performance, it plays a vital role in the field of machine learning. SVM is also immune to the issues like limitations of data dimensionality and limited samples [7,8]. When the dataset tends to infinity traditional statistical classifier provides ideal results [9], but in most real cases samples are small and limited. For realistic evaluation, we have chosen the SVM for IMF classification. Savitzky-Golay filter (SGF), also known as least-squares smoothing filter is used to smoothen noise dominant IMFs. In [11] the experimental results illustrate SGF works better in low to medium range SNR conditions

In this paper for the adaptive selection of IMFs we have proposed an effective algorithm which exploits the features of energy based threshold value and the empirical potential of SVM. Adaptively classified noise dominant IMFs are de-noised by using the SGF filter. The signal is reconstructed with both noise free and de-noised IMFs. Later comparative results are drawn from the values of SNR (Signal to Noise Ratio), PRD (Percentage Root Mean Square Difference) and MAX ERROR of input and reconstructed signal with respect to other models.

The paper is organized as follows. Section II presents our proposed model briefly. The experimental setup and result are in section III. Finally, the paper concludes in section IV.

II. PROPOSED MODEL

Fig. 1 illustrates a complete flow diagram of our proposed model, where the input signals are decomposed using EMD. The empirical decomposition results in a finite set of IMFs. A threshold value using the Pearson Correlation Coefficient value is derived from the IMFs, and then they are separated into noise free and noise dominant IMFs. Noiseless and de-noised IMFs are used to reconstruct the signal. Later we briefly describe the Pearson Correlation based classification to show a comparison of the model.

978-1-5090-5785-6/16/$31.00 ©2016 IEEE

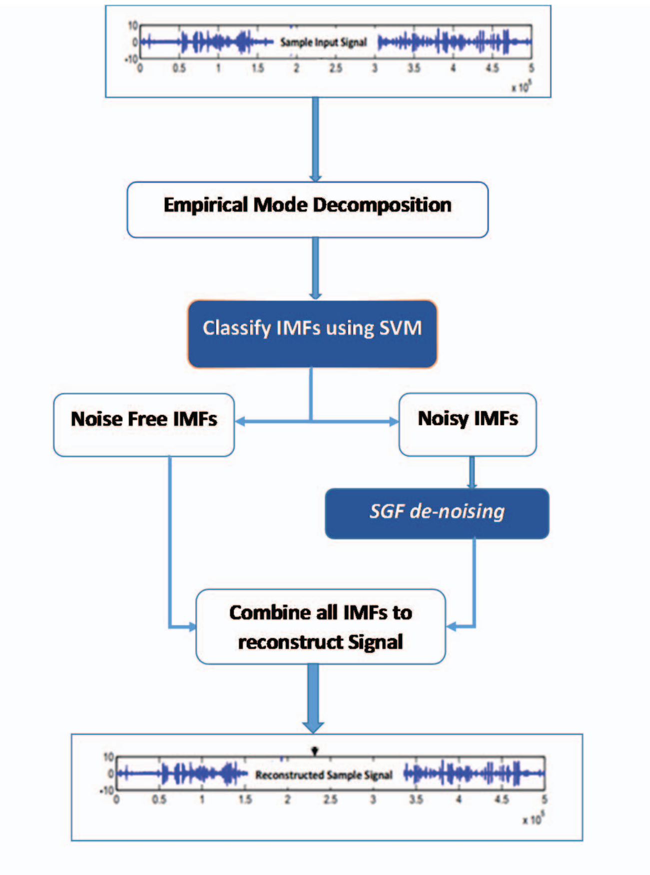

Fig. 1. Flow Diagram of the Proposed Model.

A. Empirical Decomposition and Threshold Derivation

Decomposition of the input signal using EMD results generates a finite set of IMFs. The PCC value is calculated for each of the resultant IMFs. Following the proposal of [6], a threshold value is derived (eqn.1). If Pearson's correlation coefficient of the IMFs is Pi and the threshold is expressed as T, expression of the threshold is,

$$T = \frac{\max(Pi)}{10 \times \max(Pi) - 3}\ [i = 1,2,3 \ldots] \tag{1}$$

Relevant IMFs are selected in compared to the derived T. IMF is relevant if $Pi \geq T$, irrelevant if not. This aids in training the SVM.

B. IMFs Classification using SVM

In order to train the SVM we used a subset of the decomposed IMFs. The SVM object was trained using the following features of IMFs, such as standard deviation [13] and root mean square (RMS) [12]. Then the trained SVM object is used to classify a large set of IMFs, into two categories, noise free and noisy IMFs.

C. De-noise the Noisy IMFs

Savitzky-Golay filter is used to de-noise the noisy IMFs. In our experimental setup the filter is used with frame size of 41 and 3rd order polynomial.

D. Signal Reconstruction

The filtered signal is reconstructed using noise free and de-noised IMFs.

$$S_R = \sum IMF_{nf} + \sum IMF_{dn} \tag{2}$$

where S_R represents the reconstructed signal and, IMF_{nf} and IMF_{dn} are noise free and de-noised signals respectively. We have used the original signal and the reconstructed signal to compare performance of our proposed model and other state of the art models.

III. Experimental Setup and Result

MATLAB® version 14 simulation tool is used to evaluate the performance of the proposed model. In order to validate our proposed model, we have used the dataset in [10]. We tested the dataset with a medium level randomized white noise. Performance of our proposed model and other state of art models were compared using three distinct statistical parameters, such as PRD, SNR and Max Error. Fig. 2 demonstrates a visual difference in performance of different models and our proposed model using a sample signal.

$$PRD = \sqrt{\frac{\sum_{n=0}^{N}(S(n) - S_R)}{\sum_{n=0}^{N}(S(n))^2}} \times 100\% \tag{3}$$

PRD values represent the percentage RMS difference of the reconstructed signal. The mathematical formula or PRD is as given equation 3, where $S(n)$: original signal and S_R: reconstructed signal. Table I displays calculated PRD values for the signals reconstructed using five different methods. In sense, lower PRD value represents better performance and closer resemblance to original signal. Each column represents values of respective methods headed by their names. The rows represent 5 distinct signals out of the used input signal set. The values show a clear trend of efficiency as the columns progress. Our proposed model establishes a clean dominance over other methods traditional methods compared.

$$SNR = 20 \times \log \frac{X_O}{X_N} \tag{4}$$

The mathematical equation of SNR is as given equation 4, where X_O represents the signal and X_N represents the noise added. Equation 5 is the mathematical formula for Max Error, where S_R: reconstructed signal and S_O: original signal. As presented in Table I-III our proposed model out performs other state of art models. Fig. 3 demonstrates the performance of proposed model with respect to SNR. Our proposed model gives a consistent performance over a large data set as depicted in Fig. 3, but for better readability only 5 sample signals is presented in Table II.

$$ME = |S_R - S_O| \tag{5}$$

Max Error also works as a good performance evaluating parameter as suggested in [11]. The mathematical formula for Max Error (eqn. 5) calculation is simple, S_R: reconstructed signal and S_O: original signal. The Table III shows that our proposed model achieved the most efficient performance considering Max Error values over all the others models.

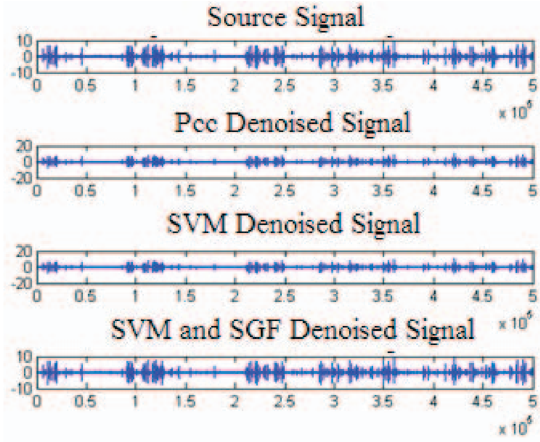

Fig. 2. Source and Reconstructed Signals.

TABLE I. CALCULATED PRD VALUES OF THE DIFFERENT METHODS

Signal	First 3 IMFs are Used in Construction	First 7 IMFs are Used in Construction	SVM Classified IMFs are Used in Construction	PCC Classified IMFs are Used in Construction	Proposed Model (SVM+SGF) Construction
1	35.0789	7.6093	16.6681	16.6681	0.8141
2	88.4465	11.0027	17.5735	17.5735	1.0142
3	102.0182	10.5388	17.7765	17.7765	1.1053
4	440.5524	16.0600	16.0570	16.0570	1.1787
5	15224.5472	38.0415	17.9300	17.9300	1.3754

a. Lower PRD value represents better performance.

TABLE II. CALCULATED SNR VALUES OF THE DIFFERENT METHODS

Signal	First 3 IMFs are Used in Construction	First 7 IMFs are Used in Construction	SVM Classified IMFs are Used in Construction	PCC Classified IMFs are Used in Construction	Proposed Model (SVM+SGF) Construction
1	9.4480	22.3739	15.5945	15.5945	41.7895
2	3.5444	19.1736	15.1356	15.1356	39.8797
3	2.9079	19.5477	15.0452	15.0452	39.1339
4	0.2411	15.9189	15.9207	15.9207	38.5749
5	0.0000	8.8114	14.9718	14.9718	37.2353

b. Higher SNR value represents better performance.

TABLE III. CALCULATED MAX ERROR VALUES OF THE DIFFERENT METHODS

Signal	First 3 IMFs are Used in Construction	First 7 IMFs are Used in Construction	SVM Classified IMFs are Used in Construction	PCC Classified IMFs are Used in Construction	Proposed Model (SVM+SGF) Construction
1	3.8788	0.7877	1.6079	1.6079	0.0862
2	6.5583	1.3310	2.7769	2.7769	0.1377
3	6.7789	1.3331	2.5055	2.5055	0.3358
4	7.4993	1.7521	1.7569	1.7569	0.0909
5	10.0021	3.9285	2.0450	2.0450	0.1432

c. Lower Max Error value represents better performance.

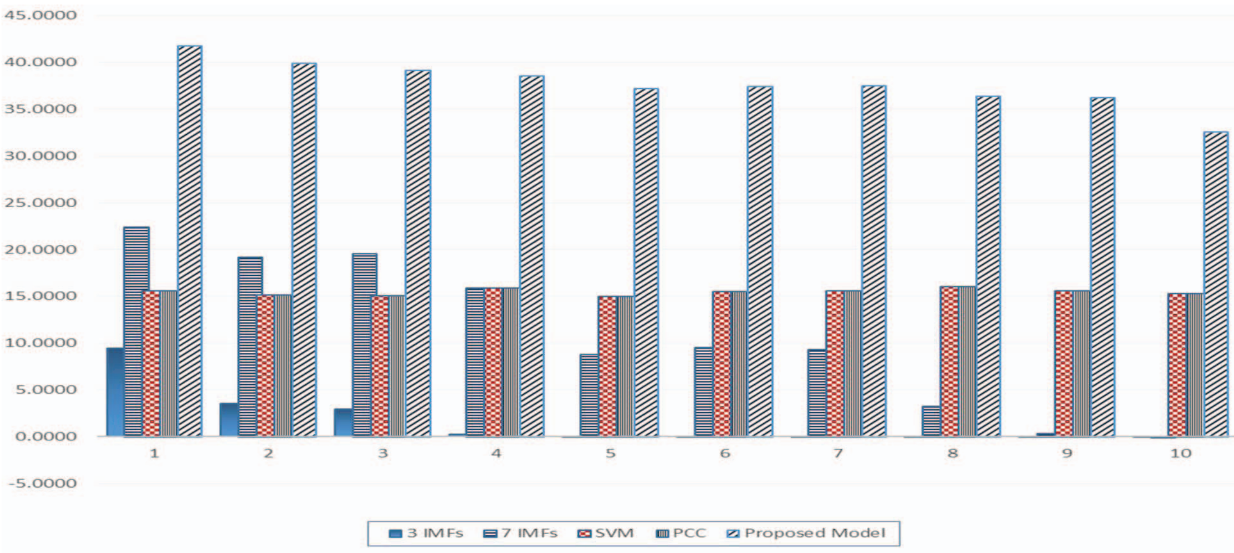

Fig. 3. SNR values for different methods.

IV. Conclusion

This paper presents a new approach to select adaptive IMF of EMD. In the proposed model we have used a machine learning tool, SVM for classification of noisy and noiseless IMFs. Savitzky-Golay filter is used to de-noise the noisy IMFs. Later the signal is reconstructed with both noise free and de-noised IMFs. Three statistical parameters SNR, PRD and Max Error were used as measuring tools in experimental evaluation. The experimental results exhibits that the proposed model out performs other state of art models by achieving a maximum 41.79% SNR, PRD and Max Error are reduced to 0.814% and 0.081% respectively, in medium noisy environments.

References

[1] S.W. Smith, "The Breadth and Depth of DSP. In The Scientist and Engineer's Guide to Digital Signal Processing," San Diego, CA: California Technical Pub, 1997, pp. 1-10.

[2] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis," in *Proc. of the Royal Society of London A:Mathematical, Physical and Engineering Sciences*, 1998, vol. 454, pp. 903–995.

[3] P. Nguyen, M. Kang, J. M. Kim, B. H. Ahn, J. M. Ha, B. K. Choi, "Robust Condition Monitoring of Rolling Element Bearings Using De-Noising and Envelope Analysis with Signal Decomposition Techniques," *Expert Systems With Applications,* 2015, vol. 42, pp. 9024–9032.

[4] D.B. Souza, J. Chanussot, and A. Favre, "On selecting relevant intrinsic mode functions in empirical mode decomposition: An energy-based approach," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 2014, pp. 325-329 .

[5] Z.K. Peng, Peter, W. Tse, and F.L. Chu, "A comparison study of improved hilbert-huang transform and wavelet transform: Application to fault diagnosis for rolling bearing," *Mechanical Systems and Signal Processing*, 2005, vol. 19, no. 5, pp. 974-988,

[6] A. Ayenu-Prah and N. Attoh-Okine, "A criterion for selecting relevant intrinsic mode functions in empirical mode decomposition," *Advances in Adaptive Data Analysis, Theory and Applications,* 2010, vol. 2, no. 1, pp. 1–24.

[7] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," In *Proc. of the Fifth Annual Workshop on Computational Learning Theory*- COLT 1992, pp. 144-152.

[8] V. Vapnik, The Nature of Statistical Learning Theory, NY: Springer-Verlag, 1995

[9] D.K. Srivastava, L. Bhambhu, "Data Classification Using Support Vector Machine," *Journal of Theoretical and Applied Information Technology*, 2010, vol 12. no. 1, pp. 435-445.

[10] J. Uddin, R. Islam, J. Kim, and C. Kim, "A Two-Dimensional Fault Diagnosis Model of Induction Motors using a Gabor Filter on Segmented Images". *International Journal of Control and Automation (IJCA*), 2015, vol. 9, no. 1, pp. 11-22.

[11] M. Almahamdy and H. B. Riley, "Performance Study of Different Denoising Methods for ECG Signals." *Procedia Computer Science,* vol. 37, pp. 325-332.

[12] M. Hamedi, S. Salleh, and A.M. Noor, "Facial neuromuscular signal classification by means of least square support vector machine for MuCI" *Applied Soft Computing*, 2015, vol. 30, pp. 83-93.

[13] P. Bhuvaneswari and J. S. Kumar, "Support Vector Machine Technique for EEG Signals," *International Journal of Computer Applications (IJCA)*, 2013, vol. 63, no. 13, pp. 1-5.

[14] V. Venkatachalam and J.L. Aravena, "Nonstationary signal enhancement using the wavelet transform," in *Proc. of the Twenty-Eighth Southeastern Symposium on in System Theory*, 1996, pp. 98–102.