

Fast and accurate primary user detection with machine learning techniques for cognitive radio networks

G. A. Pethunachiyar¹, B. Sankaragomathi²

¹Department of Computer Applications, The Tamilnadu Dr. Ambedkar Law University, India.

²Department of Electronics and Instrumentation Engineering, National Engineering College, India

Article Info

Article history:

Received Apr 3, 2020

Revised Jul 8, 2020

Accepted Aug 18, 2020

Keywords:

Decision tree

K-Nearest neighbor

Orthogonal Frequency Division

Multiplexing

Primary User

Support vector machines

ABSTRACT

Spectrum decision is an important and crucial task for the secondary user to avail the unlicensed spectrum for transmission. Managing the spectrum is an efficient one for spectrum sensing. Determining the primary user presence in the spectrum is an essential work for using the licensed spectrum of primary user. The information which lacks in managing the spectrum are the information about the primary user presence, accuracy in determining the existence of user in the spectrum, the cost for computation and difficult in finding the user in low signal-to noise ratio (SNR) values. The proposed system overcomes the above limitations. In the proposed system, the various techniques of machine learning like decision tree, support vector machines, naive bayes, ensemble based trees, nearest neighbour's and logistic regression are used for testing the algorithm. As a first step, the spectrum sensing is done in two stages with orthogonal frequency division multiplexing and energy detection algorithm at the various values of SNR. The results generated from the above algorithm is used for database generation. Next, the different machine learning techniques are trained and compared for the results produced by different algorithms with the characteristics like speed, time taken for training and accuracy in prediction. The accuracy and finding the presence of the user in the spectrum at low SNR values are achieved by all the algorithms. The computation cost of the algorithm differs from each other. Among the tested techniques, k-nearest neighbour (KNN) algorithm produces the better performance in a minimized time.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

G.A. Pethunachiyar

Department of Computer Applications

The Tamilnadu Dr. Ambedkar Law University

Chennai, Tamilnadu, India

E-mail: apethu@gmail.com

1. INTRODUCTION

In today's wireless environment, the usage of communication technologies is rapidly increasing day to day. The spectrum scarcity is increased because of the wide usage of wireless technologies. Cognitive radio (CR) Technology has the solution to the scarcity problem. The dynamic spectrum access advantages and the applications of CR technology discussed [1, 2]. There are two types of users in CR technology. The licensed user otherwise called as the primary user, the unlicensed user is known as the secondary user who can use the spectrum of a primary user without affecting the activity of the primary user. Secondary users can avail the spectrum by sensing the spectrum continuously for detecting the spectrum holes. Energy Detection algorithm is an efficient for spectrum assignment in CR. It works well in low SNR values and in the

applications that has the minimum power. The author presented the performance of the system by examining the results of the probability of detection, error statistics and the probability of false alarm [3]. CR is an intelligent communication technology in wireless environment to increase the efficiency of the spectrum by using the available bandwidth at the maximum.

The main problem of the spectrum sensing is to detect the primary user presence in the spectrum. Energy detection for the primary user detection is used for spectrum efficiency [4, 5]. To overcome the limitation of fixed threshold, spectrum sensing with the adaptive threshold is presented. Adaptive threshold detects the signal at low SNR rates [6]. The fundamental and important mechanism of cognitive radio networks (CRN) is to sense the spectrum. The author explained the spectrum sensing with additive white gaussian noise (AWGN) over the fading channels. The steps involved in setting the threshold for the spectrum sensing is analyzed [7]. The new algorithm for adaptive threshold determination is presented. The author used the Welch's and Discrete Wavelet Packet Transform energy detection methods with the adaptive threshold for detection of the user in the spectrum [8]. OFDM based energy Detection algorithm is used in detecting the presence of the user in the spectrum with the timing offset. The accurate threshold is determined without calculating the timing offset [9]. Energy detection algorithm with AWGN is added to the measured signal for transmission. It detects the signal at low SNR values. The probability of detection and false alarm are analyzed for the different SNR values and for the different number of samples [10]. The machine learning algorithms is used for improving the spectrum sensing [11-17]. Among the machine learning algorithms prescribed by the author, KNN and random forest for the spectrum usage opportunistically. These algorithms are used for detecting the 4G signal for making use of the spectrum for other applications. The ensemble algorithms with machine learning techniques are used for training and detecting the Primary user signal. Support vector machines (SVM) and neural networks are also used for analysis. The importance of cognitive radio networks and its applications in various fields are discussed by the author [18-21]. None of the above algorithms use the energy detection with machine learning for spectrum sensing. The proposed system uses the OFDM based energy detection method for spectrum sensing and machine learning. The proposed method uses the different machine learning algorithms which give the more accuracy results in comparison with other spectrum sensing techniques. The supervised machine learning algorithms contains the regression and classification algorithms. In regression, logistic regression is used for detecting the primary user and in classification algorithms SVM, decision trees, ensemble based tress, naive bayes and KNN are used.

The contributions of the proposed system are summarized as follows.

- a) The methods used in the proposed system are more auspicious than the techniques which are mentioned above in spectrum sensing.
- b) The proposed detection algorithm produces 100% accuracy in detecting the primary user signal at the licensed spectrum.
- c) This system uses the classification and regression algorithms and OFDM based energy detection signal for its detection.
- d) Data set is generated based on the proposed algorithm and it is used for training and validating the machine learning techniques.
- e) The proposed system performance is analyzed with the classification and regression technique. The results of the algorithm are compared based on the speed, accuracy time taken for training, confusion matrix and area under the curve (AUC).
- f) The performance measures based on confusion matrix also analyzed and presented for the proposed algorithms.
- g) All the detection method used in this algorithm produces 100% accuracy in detection and the performance is varied in speed and time.

The arrangement of the paper is as follows. The concept of Machine Learning is presented in Section 2, Section 3 gives the algorithm used for the data set generation, Section 4 gives the performance analysis of the proposed classification algorithms and the logistic regression algorithm and last, the paper is concluded with the conclusion and future works.

2. MACHINE LEARNING FOR SPECTRUM SENSING

Machine learning automates and analyzes the model for learning the data. It uses the artificial intelligence for learning the data, pattern identification and make decisions based on the analysis. There are many types of machine learning. Among the types, supervises learning algorithms are mostly used by the many real time applications.

Supervised learning algorithms have many inputs and single output from the inputs. Inputs can be named as Z (Z takes many inputs as Z_1, Z_2, \dots, Z_n) and learn the input values to produce the output. Here, the output is named as X and the techniques of machine learning algorithms are used to map the inputs to the output.

$$X = f(Z) \quad (1)$$

The aim of the function is to predict the output from the input variables after learning the data. The machine learning algorithms used for analysis are described as follows.

2.1. Classification algorithms

2.1.1. K-nearest neighbor algorithm

KNN is a simple and effective machine learning algorithm. The observations are classified into several categories according to the size of the variables in database. Dividing the observations into many classes are referred as a method of free distribution. It is more suitable for CR because the parameters and the size used in cognitive radio are large and it has the capability to separate data into different categories. KNN is implemented with the following steps. The first step is to transform the inputs into the vectors. The method works by measuring the distance between the vectors. Euclidean distance is the mostly used in calculating the distance. If the vectors are named as a and b where $a = \{a_1, a_2, a_3 \dots a_n\}$ and $b = \{b_1, b_2, \dots b_n\}$. The distance between the vectors are calculated using the (2).

$$d(a, b) = d(b, a) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (2)$$

The formula is used to determine the distance and to test the values in the data set. The probability of the input points are determined and similarity with the test data are identified. The classification is done according to the highest probability.

2.1.2. Support vector machines

SVM transforms the input data into the needed form. The kernel is used with the SVM for transforming the input. The low dimensional input data is converted into the high dimensional with the help of kernel. Addition of dimension for transforming the non separable into separable classes. Robustness, accuracy and flexibility is provided by SVM algorithm with the kernel. For determining the linear kernel, the dot product between each two products are calculated. The polynomial kernel is the more generalized form of the above kernel and it is used for data with more number of dimensions.

2.1.3. Naive bayes

Bayes' theorem is the basis for Naive Bayes algorithm. Here, the independent assumptions between predictors are calculated. It is chosen because of its simplicity in building the model, producing efficient results for large data-sets and its usage of parameters in iterations. This algorithm outperforms for the most of the real time applications. For any class c, the predictor x is the independent of all the predictor values. The posterior probability is determined using the following (3).

$$\frac{P\left(\frac{c}{x}\right) = P(x/c)P(c)}{P(x)} \quad (3)$$

Where $P(c|x)$, $P(c)$, $P(x|c)$ and $P(x)$ are the posterior probability, prior probability, probability of predictor and prior probability of predictor respectively.

2.1.4. Ensemble based classifier

The number of weak classifiers are combined to generate the ensemble based algorithm. The data set is trained by combining all the weak classifiers. The retraining of the classifier is avoided by updating the data in the source. By combining all the week classifiers, the prediction in accuracy is improved. Decision trees and ensembles trees are used for testing the accuracy of the results of the data set. Decision trees is the mostly used powerful and popular technique for classification in the decision tree algorithm. It is like tree structure where node, branch and leaf represents the test on the data, outcome from the tested data and class label of the each attribute.

2.1.5. Logistic regression

If the output variable is real then the regression can be used. In our proposed system, logistic regression used when the dependent variable is of the form binary. The binary data is used in the outcome predictor variable. It finds the relationship between the binary predictor variable and with all other variables.

3. PROPOSED ALGORITHM FOR DATASET GENERATION

The steps involved in generating the data set generation for analysis are as follows.

1) Selection of signal to noise ratio.

The user can select any SNR value starting from -15 dB for generating the database. Because the algorithm produced here achieves the accurate probability of detecting the primary user from any SNR value starting from -15 dB. In [22], the author presented the blind spectrum sensing method for the primary user detection. This algorithm provides the results above 90% from the SNR value - 3 dB. Compared with this, the proposed algorithm shows the better results even in SNR value -15dB.

2) Number of Iterations

The number of iterations chosen for the data set generation is 100. With the minimum number of iterations data set are generated,

3) Modulation type

The modulation type used is quadrature phase shift keying (QPSK). The modulation type can be any one of the base phase shift keying and quadrature amplitude modulation.

4) Signal generation using OFDM

- a) Signal=OFDM (Modulation Type)
- b) Calculation of the power of the signal
- c) Linear SNR calculation for the given SNR
- d) Noise Variance for OFDM signal with linear SNR

5) Signal generation using energy detection

- a) Noise generation from noise variance
- b) Signal generation for the linear SNR
- c) Add AWGN to the generated Signal
- d) Add signal and noise to transmit the signal

6) Energy and threshold determination for the generated signal

7) Compare energy and threshold for detecting the presence of primary user at the spectrum

8) Print the value of Signal, Noise, Pf, Pd, energy and threshold and concatenate all the values to make it as a Database.

9) goto Step 1 for updating the SNR value.

10) If signal is negative then it is considered as 0 for optimization that the primary user is absent in the spectrum.

11) If signal and noise is positive then the primary user is Present and it is considered as 1.

12) If noise is negative and signal is positive then considered as primary user is present and it is considered as 1

13) If both signal and noise are negative then spectrum is available for secondary user usage.

14) Database is optimized based on the steps from step10 to step 13.

4. RESULTS AND DISCUSSION

The data set is generated using the above algorithm. Data set with 5000 records are considered for obtaining the performance measures of the proposed system. Data set is divided into 70% and 30% for training and testing respectively. Table 1 describes the parameters used in the primary user prediction. For training 3500 records are used. Trained model is used for validation. Testing the performance of the trained model, 1500 records are used for validation. The parameters used in simulation in Table 1.

Table 1. Parameters used in simulation

Field Name	Field Description
Signal	The value of the signal
Noise	Noise in the signal
PF	Probability of False Alarm
PD	Probability of Detection
Energy	Energy level in the signal
Threshold	Threshold for the generated signal
Signal_Opt	Optimized Signal Value
Noise_Opt	Optimized Noise Value
PU	Primary User Presence

4.1. Confusion matrix

Confusion matrix is used as an efficient measuring parameter for getting the measured performance values of the algorithm used. Table representation of confusion matrix is described in Table 2. For training 3500 records are used. Trained model is used for validation. Testing the performance of the trained model, 1500 records are used for validation. Confusion matrix is used as an efficient measuring parameter for getting the measured performance values of the algorithm used.

Table 2. Confusion matrix

Observations	Predicted (1)	Predicted (0)
Actual (1)	TP	FN
Actual (0)	FP	TN

Where P stands for Positive and N stands for Negative

True Positive (TP): Both the observation and Predicted is True

True Negative (TN): Both the observation and Predicted is True

False Negative (FN): Predicted negative for the observation positive

False Positive (FP): Predicted positive for the observation negative

Confusion Matrix obtained for the machine learning algorithms KNN, SVM, Decision Tree, Naive Bayes and Linear Regression are depicted in Table 3. The results of the proposed system out performs all other algorithms presented in [23-25].

Table 3. Confusion matrix of decision trees, naive bayes, KNN, SVM and logistic regression

Number of Observations=1500	Predicted (1)	Predicted (0)
Actual (1)	744	0
Actual (0)	0	756

4.2. Computational cost of algorithms

The computational cost of prediction speed, training time, and accuracy are determined using different machine learning algorithms. The performance measures of all the algorithms except KNN are described in Table 4. The accuracy rate produced by all the algorithms except ensemble boosted trees is 100%. For ensemble boosted trees, it produces only 50.4 % accuracy. It fails to predict the negative results. The negative results are predicted as positive. Among all the algorithms, the KNN produces the result with 100% accuracy in minimum time. The KNN results for the different neighbour values are also considered for analysis.

Weighted KNN is an improved version of KNN. It depends of the number of neighbours. If number of neighbours is small, then it results in increasing the errors in classification result. Otherwise, there is a chance of having values from other classes in prediction. The results of KNN with different neighbour values are placed in Table 5.

Table 4. Computational cost of machine learning techniques except KNN

Name of the Algorithm	Accuracy	Prediction Speed	Training Time
Decision Trees	100%	~26000 obs/sec	4.7237 sec
Ensemble Bagged Trees	100%	~26000 obs/sec	4.7237 sec
Ensemble Boosted Trees	50.4%	~72000 obs/sec	12.22 sec
Naive Bayes	100%	~930 obs/sec	13.607 sec
SVM	100%	~19000 obs/sec	9.9632 sec
Logistic Regression	100%	~27000 obs/sec	27.124 sec

Table 5. Computational cost of weighted KNN with euclidean distance

Number of Neighbors	Distance Weight	Accuracy	Prediction Speed	Training Time
5	Squared Inverse	100%	~27000 obs/sec	2.58 sec
10	Squared Inverse	100%	~20000 obs/sec	1.5115 sec
20	Squared Inverse	100%	~11000 obs/sec	1.5769 sec
5	Inverse	100%	~29000 obs/sec	1.5684 sec
10	Inverse	100%	~27000 obs/sec	1.5648sec
20	Inverse	100%	~18000 obs/se	1.4775 sec

4.3. Area under the curve (AUC)

One of the commonly used methods of measuring the accuracy of the model is AUC. The value of AUC is from 0.5 to 1.0. 0.5 denotes the no discrimination for the predictors. If AUC lies between 0.7 and 0.8 then the model is considered as acceptable. If the values are between 0.8 and 0.9 then the model is considered as the best classifier model. If the value is above 0.9 then the performance of the model is considered as the excellent and it is considered as the correct classifier model for the corresponding data set. AUC of the proposed system is displayed in Figure 1.

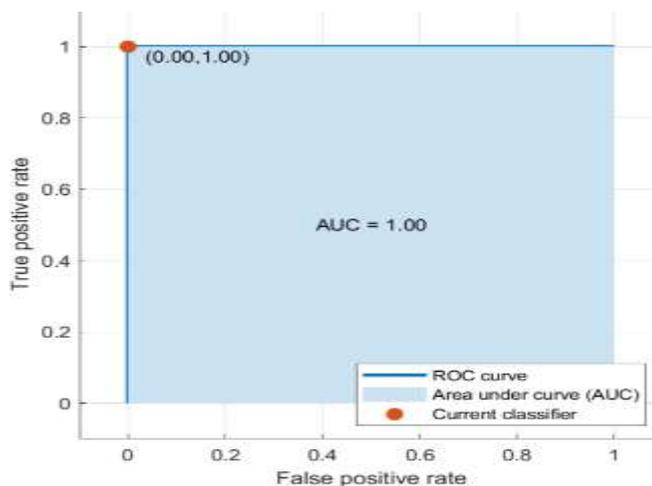


Figure 1. Area under the curve for the proposed machine learning algorithms

5. CONCLUSIONS AND FUTURE WORK

In the proposed system, the machine learning based spectrum sensing is used. It is an essential for cognitive radio networks to detect the presence of the primary user in low signal to noise ratio values. Data set is generated for the low SNR values. The data are classified according to the presence of the primary user. To train and validate the model, training and testing data set is generated from the data set. KNN, SVM, naive bayes, decision trees, ensemble based trees and logistic regression are used in analysis by extracting the features of the corresponding classifier model. Confusion matrix, computational cost and area under the curve are used to compare the performances of the algorithms. The results show that the proposed system outperforms all other algorithms in predicting the presence of the primary user. As a future work, other spectrum sensing algorithms can be investigated by extracting the more features of the signal. Here, primary user presence is considered for analysis. The detailed analysis of cognitive radio networks can be explored by including the number of primary users and secondary users in the spectrum for the better performance and for the better utilization of the spectrum.

REFERENCES

- [1] Kolodzy, "Spectrum policy task force," Rep. ET Docket 02-135, Federal Communications Commission, Washington, DC, USA, 2002.
- [2] Akyildiz MVIF, Lee WY, Mohanty S, "NeXt generation dynamic spectrum access cognitive radio wireless networks: a survey," *Computer Networks Journal* vol. 50, no. 13, pp. 2127-2159, 2006.
- [3] Ghosh, S.K., Mehedi, J. & Samal, U.C, "Sensing performance of energy detector in cognitive radio Networks," *International journal of information Technology*, vol. 11, no. 4, pp. 773-778, 2019.
- [4] Patel, *et al*, "Spectrum Sensing with Energy Detection in Cognitive Radio Networks," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 3, pp. 1301-1308, 2017.
- [5] P. Pandya, A. Durvesh and N. Parekh, "Energy detection based spectrum sensing for cognitive radio network," *2015 Fifth International Conference on Communication Systems and Network Technologies*, Gwalior, 2015, pp. 201-206.
- [6] M. Z. Alom, T. K. Godder, M. N. Morshed and A. Maali, "Enhanced spectrum sensing based on Energy detection in cognitive radio network using adaptive threshold," *2017 International Conference on Networking, Systems and Security (NSysS)*, pp. 138-143, 2017.
- [7] Mahmood A., *et al*, "Energy detection technique for spectrum sensing in cognitive radio: A survey," *International Journal of Computer Networks & Communications (IJCNC)*, vol.4, no.5, pp. 223-242, 2012

- [8] Wang, Y. Gao and L. Cuthbert, "Spectrum sensing using adaptive threshold based energy detection for OFDM signals," *2014 IEEE International Conference on Communication Systems*, pp. 359-363, 2014.
- [9] Shi, Z., *et al.*, "Improved spectrum sensing for OFDM cognitive radio in the presence of timing offset," *Journal of Wireless Communications Networks*, vol. 2014, p. 224, 2014,
- [10] Pethunachiyar G.A & Sangaragomathi B, "An Improved Energy Detection algorithm for Sepctrum Sensing in Cogniitve Radio Networks," *International Journal of analytical and experimental modal analysis*,2020
- [11] O. P. Awe, Z. Zhu and S. Lambotharan, "Eigenvalue and Support Vector Machine Techniques for Spectrum Sensing in Cognitive Radio Networks," *2013 Conference on Technologies and Applications of Artificial Intelligence*, Taipei, 2013, pp. 223-227.
- [12] Wasilewska, M., & Bogucka, H., "Machine Learning for LTE Energy Detection Performance Improvement," *Sensors*, vol. 19, no. 19, p. 4348,2019
- [13] D. Zhang and X. Zhai, "SVM-Based spectrum sensing in cognitive radio," *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*, Wuhan, 2011, pp. 1-4.
- [14] Y. Tang, Q. Zhang and W. Lin, "Artificial Neural Network Based Spectrum Sensing Method for Cognitive Radio," *2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, Chengdu, 2010, pp. 1-4.
- [15] Feng Q, Y Zhihui and S Keqin, "Spectrum environment machine learning in cognitive radio," *Procedia Engineering*, vol. 29, pp. 4181-4185, 2012.
- [16] Haozhou Xue and Feifei Gao, "A machine learning based spectrum-sensing algorithm using sample covariance matrix," *2015 10th International Conference on Communications and Networking in China (ChinaCom)*, Shanghai, 2015, pp. 476-480.
- [17] Hassaan Bin Ahmad, "Ensemble classifier based spectrum sensing in cognitive radio networks," *Wireless Communications and Mobile Computing*, vol. 2019, p. 9250562, 2019.
- [18] Miguel Tuberquia, Hans Lopez-Chavez, Cesar Hernandez, "Bio-inspired route estimation in cognitive radio networks," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, pp. 3095-3107, 2020.
- [19] Thi-Anh Hoang, *et al.*, "Security performance analysis for power domain NOMA employing in cognitive radio networks," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 3, pp. 1046-1054, 2020.
- [20] Ankur S Kang, Renu Vig, "K coverage probability of 5G wireless cognitive radio network under shadow fading effects," *Indonesian Journal of Electrical Engineering and Informatics*, vol. 4, no. 3, pp. 181-188, 2016.
- [21] Ali Suzain, *et al.*, "Machine learning based lightweight interference mitigation scheme for wireless sensor network," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 18, no. 4, pp. 1762-1770, 2020
- [22] Kai Yang, Zhitao Huang, Xiang Wang, and Xueqiong Li, "A Blind Spectrum Sensing Method Based on Deep Learning," *Sensors*, vol.19, no. 10, p. 2270, 2019.
- [23] Tian, Jinfeng, *et al.*, "A machine learning-enabled spectrum sensing method for OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11374-11378, 2019.
- [24] J. Gao, X. Yi, C. Zhong, X. Chen and Z. Zhang, "Deep Learning for Spectrum Sensing," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1727-1730, 2019.
- [25] Hurmat Ali Shah and Insoo Koo,"Reliable Machine Learning Based Spectrum Sensing in Cognitive Radio Networks," *Wireless Communications and Mobile Computing*, vol. 2018, p. 5906097, 2018.