

Applying Classification Methods for Spectrum Sensing in Cognitive Radio Networks: An Empirical Study

Nayan Basumatary, Nityananda Sarma and Bhabesh Nath

Abstract Spectrum sensing is the paramount aspect of cognitive radio network where a secondary user is able to utilize the idle channels of the licensed spectrum band in an opportunistic manner without interfering the primary (license) users. The channel (band) is considered to be idle (free) when primary signal is absent. The channel accessibility (free) and non-accessibility (occupied) can be modeled as a classification problem where classification techniques can determine the status of the channel. In this work supervised learning techniques is employed for classification on the real-time spectrum sensing data collected in test bed. The power and signal-to-noise ratio (SNR) levels measured at the independent CR device in our test bed are treated as the features. The classifiers construct its learning model and give a channel decision to be free or occupied for unlabelled test instances. The different classification technique's performances are evaluated in terms of average training time, classification time, and F1 measure. Our empirical study clearly reveals that supervised learning gives a high classification accuracy by detecting low-amplitude signal in a noisy environment.

Keywords Cognitive radio · Spectrum sensing · Primary user detection
Supervised learning techniques

N. Basumatary (✉) · N. Sarma · B. Nath
Department of Computer Science and Engineering, Tezpur University,
Napam, Assam, India
e-mail: nayanb@tezu.ernet.in

N. Sarma
e-mail: nsarma@tezu.ernet.in

B. Nath
e-mail: bnath@tezu.ernet.in

1 Introduction

Cognitive Radio (CR) is the emerging technology in the domain of new age wireless communication. It can dynamically change its transmission parameters based on changes in environmental factors [1]. In cognitive radio network a secondary or unlicensed user can sense the licensed channels for any opportunity to transmit, which results to efficiently utilize the available channel of primary licensed users. For performing this spectrum access in an opportunistic manner the CR devices need to sense the radio spectrum licensed to primary users. So, efficient spectrum sensing is very important for opportunistic spectrum access. In Cognitive Radio [2, 3] the spectrum sensing is carried out in a co-operative and independent manner. In co-operative sensing all the CR devices co-operate with each other to take a collective decision which results into get high sensing reliability. While, in case of independent sensing each CR device performs the sensing individually and make its own sensing decision to use unoccupied spectrum portion. Here, in this work analysis of the prominent supervised learning techniques [4–6] was done for noncooperating spectrum sensing framework to decide the presence or absence of primary user in a channel.

In low SNR environment (fading channels) where there is high noise level and regardless of the fact that there is a signal present (low amplitude) it cannot be distinguished. This work exploits the signal power and the SNR feature to take a decision in such condition. The conventional energy detection method may cause misdetection of the signal as it fails in a low SNR environment. The motivation of using supervised learning [7] is that in supervised learning the classifier learns from some objects which are having some class marks and when unknown object's class is to be predicted this class marks are assigned to unknown objects based on learning done previously which actually gives high detection accuracy. Here all supervised learning models are built not only based on just the power received of the signal, but also the SNR feature so that even if there is a low power signal in a highly noisy environment the classifier can still give a decision to detect the signal with a priori knowledge.

Contribution:

1. The customized dataset was created by capturing both the power and SNR features in USRP-based test bed for performing the classification task.
2. The SNR or the Signal-to-Noise ratio value used for classification for detection of signal in low SNR environment which is not explored in literature. The classifier decision is based on SNR parameter value which helps in detection of low amplitude signal.
3. Due to high prediction accuracy analysis was carried out using known supervised learning techniques such as the SVM (Linear, Poly, and RBF kernels), Logistic regression, K-Nearest Neighbor, Gaussian Naives Bayes, and Decision Trees.
4. The performance of the classification techniques was evaluated in terms of training time, classification delay, and F1 measure.

The remaining sections of the paper are as follows. In Sect. 2 the system model and the assumptions are presented. Then the supervised learning techniques in a noncooperative framework are discussed in Sect. 3. Section 4 describes the experimental test bed setup to amass the sensing data and prepare the data set. The numerical results and plots are given in Sect. 5. Finally, Sect. 6 gives the conclusion and future work to be done.

2 Framework Model and Presumptions

Here, a homogeneous Cognitive Radio network is assumed which consist of both the Primary and Secondary Network which is formed by ' N ' number of primary users denoted as $\{PU_1, PU_2 \dots PU_N\}$. The primary users have a set of channels $C = \{1 \dots L\}$ that are used for its transmission. The channel occupancy states are assumed to be independent and any channel can be used by the primary user. The secondary network consist of a set of ' K ' users called secondary users (SU) denoted by $\{SU_1, SU_2 \dots SU_K\}$ are also present in the CR network which are sensing these primary channels to access the channels in an opportunistic way. These users are equipped with some learning techniques to detect whether the channels are free and then to access these primary channels for its own transmission. In this secondary network each user independently searches for a spectrum opportunity. Also, it was assumed that each secondary user can sense only one channel at a time and they do not co-operate with each other. Both the primary and the secondary network are assumed to be synchronized with a global clock, and transmission is done in a time-slotted manner.

The spectrum sensing in Cognitive Radio essentially decides between two states of the channel whether there is signal present or absent. The features extracted from the samples received of the signal can be analyzed to decide between two hypotheses:

$$\begin{aligned} H_0 : x(i) &= n(i) && \text{(Signal is absent.)} \\ H_1 : x(i) &= s(i) + n(i) && \text{(Signal is present.)} \end{aligned}$$

The received samples of the signal represented by $x(i)$, the primary user signal $s(i)$ and the noise $n(i)$, where i denotes the i th sample. The signal transmission and the signal received are done in a continuous way. In this paper the signal is treated as discrete since the receiver or the sensor takes discrete samples of the signal.

Here, in case of indoor environment the primary transmission is in Line-of-sight propagation (LOS) between the transmitter antenna and the receiver antenna, hence the probability density function of the fast varying amplitude of the received instantaneous signal can be described by Rician distribution. The environmental noise is assumed to be AWGN (Additive white Gaussian Noise). The average path loss $PL(d)$ for a transmitter and receiver with separation distance d is

$$PL(d) = PL(d_0) + 10n \log(d/d_0),$$

where ‘ n ’ is path loss component and indicates the rate at which path loss increases with distance d , with close in reference distance d_0 .

In a Machine Learning framework the features extracted from the signal can be used for the feature vector and then build the model for both the training and testing phases of different classifiers. The performance of the classifier can then be separated in two hypotheses:

$$H_0 = \text{Incorrect Signal Classification.}$$

$$H_1 = \text{Correct Signal Classification.}$$

3 Supervised Learning for Spectrum Sensing

Here, some of known supervised learning techniques are applied for noncooperative spectrum sensing data in cognitive radio network. As known that CR devices have cognition capability to learn from the environment, supervised learning can be an effective way to extract information from the signals present or absent in the environment at the physical level and utilize this cognizance to make some decision in the upper hierarchy levels. The Support Vector machines (SVM), Logistic Regression (LR), K-nearest Neighbor (KNN), Decision Trees are the main classifiers used in a Nonco-operating Spectrum Sensing environment in CR Networks to detect the presence of the Primary User signal based on features of the collected data. Here, the two features power and SNR value forms an object $y^{(i)}$ and which is labeled with the corresponding channel availability CA_i .

In case of Supervised learning there are two phases, training and testing where the goal is to construct a classifier to map the samples received by the secondary users to that of the labeled samples.

1. *Training*: Let the training object is of the form $y^{(i)} = (P_i, SNR_i, CA_i)$ where P_i denotes the power value, SNR_i is the Signal to Noise Ratio and CA_i is corresponding labeled channel availability status in different time slots T_i . In the training phase this objects are fed to the classifier to build its model.
2. *Testing*: After the classifier is trained it is then ready to classify the test object of the form $y^{(i)} = (P_i, SNR_i)$ and then give the corresponding class label based on its model.

Therefore, to implement supervised learning in a Non Co-operating sensing environment, the classifier should be informed about the channel status information for some object values for the purpose of training.

4 Experimental Setup

In a cognitive radio network (CRN) in which SU can access free channel/band when unutilized by Primary users (PU). The CR devices in the network are the secondary users who are sensing the primary channels in a noncooperative manner and making its own individual decision. The secondary users are equipped with a classification framework. The sensing data obtained by the secondary users are fed to the classification framework which helps the secondary user to determine whether the channel is free or not. The experimental setup for real time dataset generation is implemented with the help of GNU Radio and USRP devices [8–11]. In the setup, three PC designated as PC1, PC2 and PC3 connected with USRP1 with a RFX2400 daughter board by a USB cable. PC1 acts the transmitter and PC2 and PC3 are the receiver/sensing nodes (Fig. 1).

In GNU Radio the existing sample python scripts *usrp_spectrum_sense.py* and *benchmark_tx.py* was modified as *sensing.py* and *transmission.py* for sensing and transmission respectively.

Other program parameters in the transmission.py script are

The Sampling rate	1 Mega Samples
Modulation used	GMSK
Sub Channel Bandwidth	6.25 kHz
So, the no of fft bins collected in a particular Channel	160 (1MS/6.25e3)

The receiver which is tuned to the center frequency of the channel can sweep only 8 MHz channel Bandwidth due to the USRP1 daughter board constraint. Out of the total 160 bins 75% is taken and 25% is discarded from both the lower and upper cut frequency (12.5% each) of the channel. The program senses the power level from bin 20 to bin 140.

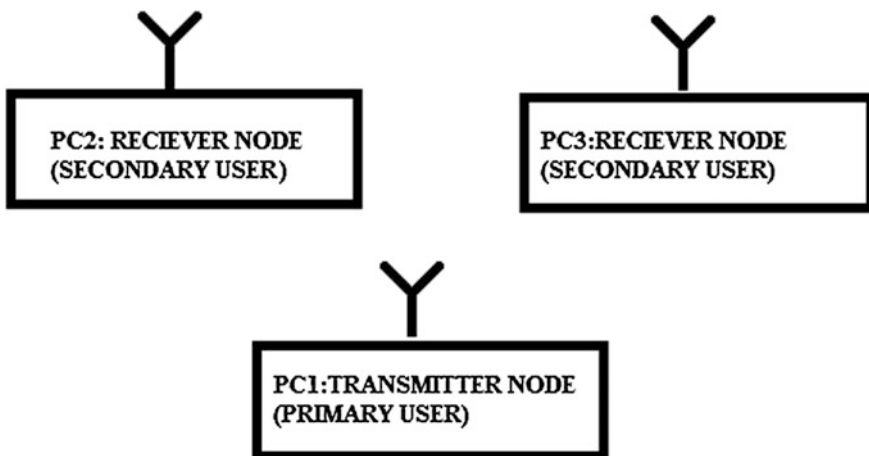


Fig. 1 Block diagram of the experimental setup

The sensing data was captured with the power and SNR features with active transmission and another with no transmission. All the sensing data are labeled as “**Free**” and “**Occupied**” class with respect to the known occupied and free channels respectively.

5 Numerical Results and Discussion

For carrying out the analysis the machine learning tool in python scikit learn [11] was used for performing the classification of the sensing data. Different Supervised Classification Algorithms that were used are Support Vector Machines with different kernels like Linear, Polynomial and RBF kernel, K-Nearest Neighbor (KNN), Logistic Regression (LR), Gaussian Naïve Baye’s (GNB) and Decision Trees (DT).

The total of samples collected from the test bed experiments was 600 after preprocessing. Cross validation random split into train and test subsets. Some percentage of the data is considered for training the classifier and the rest for testing purpose in an offline mode.

Table 1 shows the comparison of the average training time taken by different classifiers. The training time measured is high in case of SVM classifier with radial basis function (RBF) with the increase in percentage of the training samples. The Logistic Regression classifier also takes more time compared with the others.

Table 2 shows the comparison of the average classification time taken by the supervised classifiers, where the SVM classifier with radial basis function

Table 1 Average training duration of the supervised classifiers (seconds)

Training samples (%)	SVM (linear)	SVM (RBF)	SVM (polynomial-degree 3)	KNN	LR	Gaussia naïve bayes	Decision trees
50	0.001	0.004	0.001	0.001	0.001	0.001	0.001
60	0.001	0.005	0.001	0.001	0.001	0.001	0.001
70	0.001	0.006	0.001	0.001	0.001	0.001	0.001
80	0.001	0.008	0.001	0.001	0.003	0.001	0.001

Table 2 Average classification duration of the supervised classifiers (seconds)

Test samples (%)	SVM (linear)	SVM (RBF)	SVM (polynomial-degree 3)	KNN	LR	Gaussia naïve bayes	Decision trees
20	0.001	0.001	0.001	0.001	0.001	0.001	0.001
30	0.001	0.001	0.001	0.001	0.003	0.001	0.001
40	0.001	0.002	0.001	0.001	0.002	0.001	0.001
50	0.001	0.002	0.001	0.001	0.001	0.001	0.001

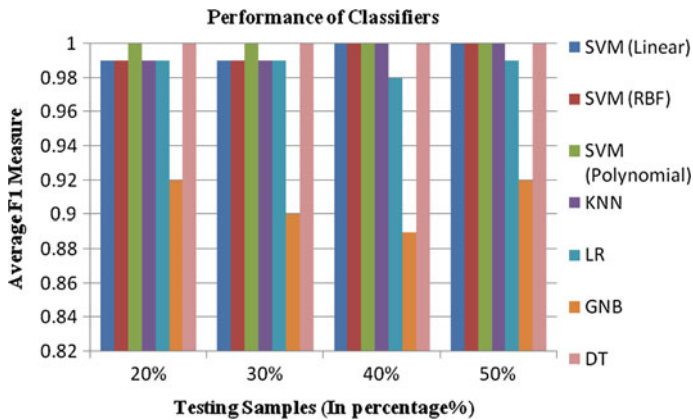


Fig. 2 Performance of classifiers with different number of testing samples and average F1 measure

(RBF) relatively takes more classification time than the other two kernel function. With the increase in percentage of the test samples the RBF classification time also increases than the others. It shows the comparison of the average classification time taken by the rest of the classifiers taken into consideration. The Logistic Regression classifier takes more time among others which are more or less having the same average time.

Figure 2 show performance of all classifiers used based on average F1 measure versus on the different percentage of test samples used for the classification. From the plot it is very clear that for all the different percentage of test samples used the SVM with polynomial kernel and decision tree performs very well than the rest of the classifiers which gives a higher primary user detection rate. The results also reveal that with the increase of percentage of the test samples the other SVM kernel functions along with KNN classifier performs well.

6 Conclusion and Future Work

In this work emphasis was on detection of primary user signal in a low SNR environment. An analysis was conducted using supervised learning techniques in a Noncooperating sensing manner based on real-time sensing data. The classifiers were trained in an offline manner to build its model and then decision for signal present or absent was done with different percentages of test samples. Compared to all the classifiers the SVM classifier with Polynomial and Decision Tree classifiers outperforms other techniques. The performance of all the classifiers was analyzed in terms of high detection rate like training duration and classification delay by using real-time dataset. The other classifiers may perform well with more number of samples in the dataset which may increase the sensing time of the secondary users.

But in a CR network if the sensing time is less to give a near-optimal detection rate of the primary user signal which is an additional advantage for secondary user communication (transmission time).

Future work could be to apply these learning techniques for cooperative spectrum sensing in cognitive radio network. This work is offline training and testing. If this can be extended to be implemented in an online fashion in real time it will be beneficial in a dynamic Cognitive Radio Network.

References

1. Haykin, S.: Cognitive radio: brain-empowered wireless communications. *IEEE J. Select. Areas Commun.* **23**, 201–220 (2005)
2. Urkowitz, H.: Energy detection of unknown deterministic signals. In: *Proceedings of IEEE*, vol. 55, pp. 523–231 April 1967
3. Cabric, S.D., Mishra, S.M., Brodersen, R.W.: Implementation issues in spectrum sensing for cognitive radios. In: *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, vol. 1, pp. 772–776, 7–10 Nov (2004)
4. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn. J.* **20**(3), 273–297 (1995)
5. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*, 2nd edn. Wiley, New York (2001)
6. Thilina, K.M., Choi, K.W., Saquib, N., Hossain, E.: Machine learning techniques for cooperative spectrum sensing in cognitive radio networks. *IEEE J. Sel. areas commun.* **31**(11) 2013
7. Kassiny, M.B., Li, Y., Jayaweera, S.K., A survey on machine Learning techniques in cognitive radios. *IEEE Commun. Surv. Tutorials* **15**(3), 1136–1159 2013
8. Software Defined Radio Forum.: www.sdrforum.org
9. Ettus Research LLC.: <http://www.ettus.com/>
10. <http://www.gnu.org/software/gnuradio>
11. <http://scikit-learn.org/>