Machine Learning in estimating Spectrum Occupancy

Shubham Khanna, 2015179[1]; Shubhankar R Butta, 2015180[1]
[1]Indraprastha Institute of Information Technology, Delhi

Abstract

We aim to analyze the spectrum occupancy in cognitive radio networks(CRNs) using different machine learning techniques.

We will mainly attempt Naive Bayesian Classifier[NBC], Support Vector Machine[SVM], Linear Regression[LR] and review HMM[hidden Markov model] to find the best classification technique on the basis of highest mean classification accuracy(CA) on the provided dataset.

We will review how the classified occupancy status can be used to determine the blocking probability of a secondary user for future time slots which can be used to design better cognitive systems.

Introduction

Cognitive Radio is an intelligent technology which can detect available/occupied spectrum channels and can change transmission parameters accordingly for smarter transmission. It detects empty spectrum regions which can be utilized for communication without interfering with other users.
In modern wireless communication, there is no idle frequency spectrum capacity. The bands which are the most suitable are already allocated or sold out. Thus, new incoming wireless systems have to be either moved to higher frequency bands or have to share the spectrum with these existing allocated systems.
As we mentioned above, the CR system manages dynamic environment of the radio spectrum. Thus, knowledge of statistics or different models for detecting spectrum occupancy will greatly benefit the building of new CR systems as well as in solving CR problems, thereby, improving its performances.
Machine learning is a science which lets computers act on something that they're not programmed for. For wireless communication systems, learning is used to solve the problems related to Cognitive Radios e.g. dynamic spectrum sharing. We will try different Machine Learning models to find channels best for transmission.

Applied Techniques

The learning classifier should allow the CR system to adapt to changing environment. Thus, we applied Machine Learning techniques like regression and Support Vector Machines (SVM) to classify whether the spectrum is occupied or not.
We found the CR data from one of the research projects of University of Washington. The data contained 24 files, 6 files of each channel receivers recorded at 4 different times of the day. This gave us a variation in time, frequency as well as location (different channel receivers are placed at different locations).
We divided our data into training and testing in the ratio 3:1. Then, we randomized the data to remove any bias from it. Then we applied k-fold technique on the cross-validation set to remove bias from the weights obtained. On applying gradient descent, we found out our optimum weights. Using these, we compared the obtained result to the "*Pij*" files provided on the same website.
We, next tried developing an algorithm which would efficiently detect spectrum occupancy for each sub-channel using grouping of bins in the channel recordings. We assumed that, if a sub-channel of bandwidth W consists of *N* bins and any one out of these *N* bins has power > *threshold :-* that sub-channel of the channel is occupied(Hence, classified – 1), otherwise not occupied. We then averaged out these per sub-channel labels(in 0s and 1s) over total 2000 time-sweeps to get the occupancy probability of the channel across sub-channels and bins.

Results

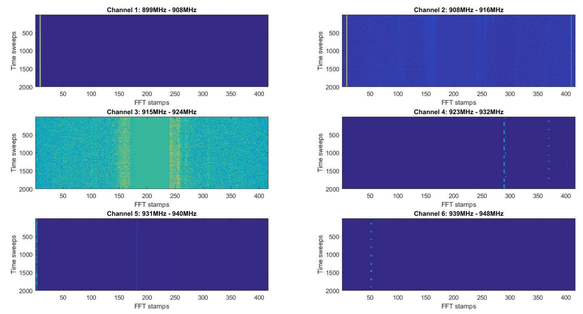
We tried applying Support Vector Machines model using Stochastic Gradient Descent on the feature set we created using the data. We compared intra-channel power outputs at different time intervals and observed the following:

- ❑ We can see that in most cases, more occupancy is observed during the day time (12 PM and 3 PM) as compared to early morning (7 AM) and night time (12:01 AM). This might be because more people are active during this time of the day as compared to late night or early morning.
- ❑ Just as predicted, machine learning techniques like logistic regression and Support Vector Machines happen to give better results than the weak models like statistical model and CDF/PDF methods. This is because of their inability to showcase the time, frequency or location variance and that they are incomplete models.
- ❑ Another common observation is that channel 4 has higher occupancy than all the other channels.

Table 1. Dataset characteristics

Property	Value
B/W / channel	8.33MHz
Channels / recording time	6
Time between rows	0.0599s
Total time / channel	119.81s
Total time sweeps	2000
Frequency / bin	20KHz

Plot 1. Spectrogram of channel data



Discussion

We are using machine learning techniques like regression and support vector machines in order to classify whether a channel is occupied or not.

ML techniques have an advantage over other methods like Statistics, CDF/PDF and Markov chain methods. These methods are comparatively simple to implement but they are either incomplete or miss out on time, frequency, location or threshold variance.
Whereas, techniques like Regression and Support Vector Machines (SVM) takes all of this into consideration, making it a better choice of model for this classification

Conclusions

We found a few limitations to our ML approach due to the lack of a suitable dataset where we could actually obtain different features like *Time of the Day*, Frequency band, Location, History of events etc.

Such a dataset, though very intuitive, was a bit tough to find for us. We settled for a CR Project by the UW[7] Which didn't contain the above features but a decent number of points to at least think about ML in spectrum occupancy.

We concluded that with better features and datasets, ML could be applied on a better dataset to get the best out of Cognitive Radio systems.

Future Directions

Machine Learning in CR system is still an emerging application of ML with great potential for the fututе. With the advent of huge number of IoT devices, increasing network traffic, a dynamic smart and *learnt* CR system would be the need of the hour. With good analysis of ML algorithms and fine parameter tuning, ML will find a huge importance in CR's future to improve its performance.

Contact Information

Shubham Khanna, Shubhankar R Butta
Indraprastha Institute of Information Technology, Delhi
Email: shubham15179@iiitd.ac.in,shubhankar15180@iiitd.ac.in
Website: iiitd.ac.in
Phone: 9891200414,9891892917

References

1.http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6933906&tag=1
2.http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6336689&tag=1
3.http://ieeexplore.ieee.org/document/7460899/?part=1
4.http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4549771
5.http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4562504
6.http://ieeexplore.ieee.org/document/7289481/references
7.http://www2.ee.washington.edu/research/funlab/funlab_CRdata/

Acknowledgements

We would like to thank Vivek sir, Parag sir and Gaurav sir for their guidance throughout this project. Parag sir helped us understand the scope of the project, Vivek sir helped us pick out the dataset and Gaurav sir helped us in analyzing and visualizing the dataset.