Advances in Intelligent Systems Research, volume 161
International Conference on Transportation & Logistics, Information & Communication, Smart City (TLICSC 2018)

Research of Radio Frequency Channel Occupancy Prediction Based on Decision Tree

Fuyin Zhao [a], Yuqi Zeng [b], Xueling Zhang [c] and Kai Zhou [d]

The State Radio Monitoring Center, Beijing 100037, China.

[a]zhaofuyin@srrc.org.cn, [b]zengyuqi@srrc.org.cn, [c]zhangxueling@srrc.org.cn, [d]zhoukai@srrc.org.cn

Abstract. Efficient utilization of finite spectrum resource is one of the purposes of spectrum management. Frequency channel occupancy is often used to evaluate the historical usage status of a frequency. Predicting the frequency channel occupancy is of great value to improve the utilization rate of spectrum resources. In this paper, an approach of extracting features from historical frequency channel occupancy data is proposed, and a method of frequency channel occupancy prediction based on decision tree is designed. Based on the actual spectrum monitoring data, we use big data and machine learning technology to predict the frequency channel occupancy of various radio services. The actual test results show that the prediction method is of high accuracy.

Keywords: spectrum management, frequency channel occupancy, decision tree, prediction, machine learning.

1. Introduction

With the rapid development of wireless communication technology, such as 5G, the Internet of things, and with the wide application of various radio services, radio spectrum resource management faces huge challenges: on the one hand, appropriate spectrum should be planned out from the rare spectrum resources for the use of new radio services, on the other hand harmful interference should be prevented to safeguard normal radio services order. Spectrum monitoring ensures the efficient management of electromagnetic spectrum by monitoring and measuring air wave parameters [1]. Current spectrum monitoring is mostly based on historical and current monitoring data, and it is impossible to predict the future usage of radio spectrum, which is of great significance to spectrum resource management.

At home and abroad, two parameters, frequency channel occupancy and frequency band occupancy are mainly used to analyze the utilization efficiency of spectrum in both the time domain and frequency domain [2]. The radio monitoring technology department have collected a large amount of monitoring data, which lays a data foundation for studying and predicting the future usage of frequency. This paper selects the frequency channel occupancy data and introduces the classic decision tree algorithm in machine learning to realize the prediction perception of frequency channel occupancy.

In the radio field, decision tree algorithm is mainly used for modulation recognition and direction finding. Reference [3] uses decision tree classifier. When the signal-noise ratio is higher than 10dB, the average modulation recognition accuracy rate of signal set {AM, FM, USB, LSB, DSB, BPSK, QPSK, OQPSK, 2FSK, 4FSK, 16QAM and 32QAM} was not lower than 93%. Reference [4] uses decision tree to predict wave coming direction all-round; when the error is less than 3 degree its accuracy can reach 91.26%, and when the error is less than 5-degree accuracy can reach 98.36%. In addition, the reasoning method and framework of spectrum management policy based on decision tree are proposed in reference [5]. When the cognitive radio equipment conducts radio-frequency emission, the decision tree is used to determine whether the radio frequency parameters conform to the relevant spectrum management policies. In this paper, a new frequency channel occupancy prediction method based on decision tree is proposed, in which the thought of feature extraction idea is applied, and parameter optimization process of decision tree algorithm is established. Here we take the typical radio services as an example to test the accuracy of the prediction method.

2. The Overall Structure of the Prediction

Machine Learning (ML) is an algorithm that uses historical data for training, automatically analyzes and obtains patterns from the data, and produces a well-trained model. When new data is provided, the model can be used for prediction. The huge volume of original spectrum data produced by spectrum monitoring activities, the frequency channel occupancy data after statistical analysis and other historical data are the data source of machine learning. Due to its huge volume and continuous increment, this paper uses Hadoop big data platform [6] to store certain historical data. Considering the efficiency of calculation, a frequency occupancy prediction method based on the classic decision tree in Spark's machine learning library (MLlib) is proposed. The overall structure of the frequency channel occupancy prediction process designed in this paper is shown in figure 1.

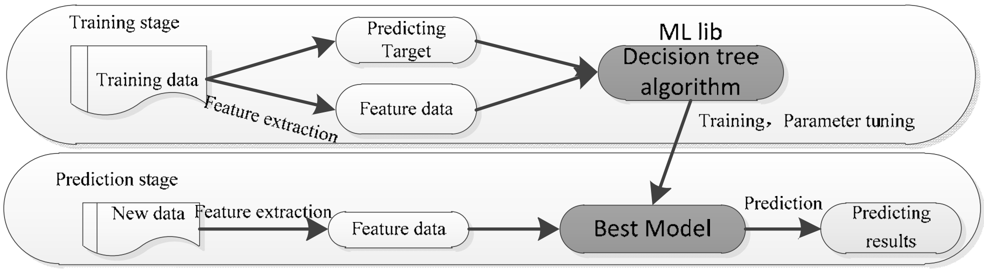


Fig. 1 Structure diagram of frequency channel occupancy prediction process

The process of frequency channel occupancy prediction is divided into the stages of training and prediction. In the training stage, feature extraction is first carried out to generate feature data and data samples of prediction targets. The data sample is used as the input data of decision tree model training, and then the decision tree prediction model is evaluated to generate the optimal decision tree prediction model. In the training stage, extracting features based on historical monitoring data and optimizing parameters of decision tree model are the most critical steps, which have a great impact on the accuracy of prediction results. After entering the prediction stage, the new feature data can be extracted and produced according to the newly generated spectrum monitoring data, and then the optimal model generated in the training stage can be used to predict and generate the prediction results.

3. Key Techniques of Research

3.1 Parameter Tuning of Decision Tree

Decision tree learning usually adopts heuristic method to induce a group of splitting rules from the training data set. When feature selection is conducted for each splitting, local optimal features are found by calculating the purity in the post-splitting node data set. The purer the node data set is, the more orderly the split rule is. In the decision tree of Spark MLlib, the ordinal parameter to estimate the node data set is impurity, whose value can be entropy, gini coefficient or variance. Entropy and gini coefficient are for classification, and variance is for regression. The frequency channel occupancy prediction in this paper uses a regression decision tree, and the value of impurity parameter is variance. The definition of variance is as follows:

$$\partial^2 = \frac{1}{\mathrm{N}} \sum_{i=1}^{N} (y_i - \mu)^2 \tag{1}$$

In the formula, y_i is the target tag of an example in the data sample, N is the total number of examples in the data sample, and μ is the mean value of target tags of all examples in the data sample.

For the feature selected for each split, if the value of the feature is too large, it is necessary to control the number of nodes after the split, which can greatly improve the training speed, but will have a certain impact on the prediction accuracy of the decision tree. The maxBins parameter is used in the decision tree of Spark MLlib to control the number of nodes after each split.

Pruning should be considered to prevent overfitting in the process of constructing decision tree. The maxDepth parameter can be used in the decision tree of Spark MLlib to control the maximum height of the tree. When the height of the tree exceeds this parameter, the split should be stopped. The higher the decision tree, the higher accuracy it will reach in the meantime the higher the training costs will be and the more easily it will overfit.

It can be seen that maxBins and maxDepth are the key parameters to determine the training cost and accuracy of the decision tree. The tuning process of these two parameters in this paper is shown in figure 2.

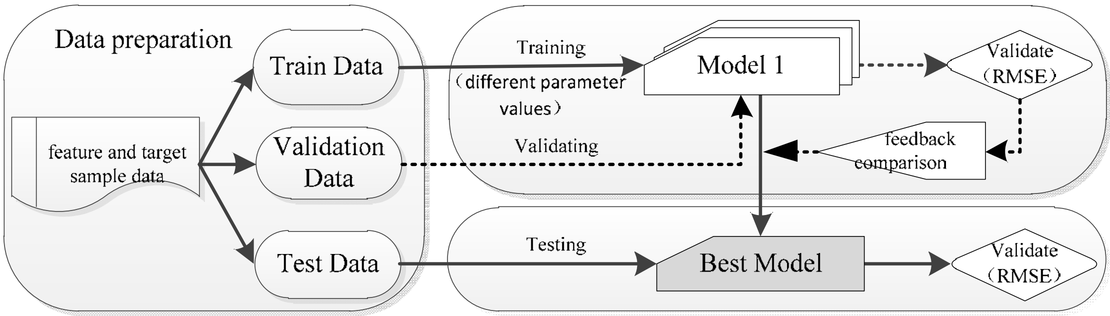


Fig.2 Parameter tuning process

Parameter tuning can be divided into three stages: data preparation, training evaluation and testing. In data preparation stage, the feature and target sample data are randomly divided into three parts: training data, evaluation data and test data. In the training evaluation stage, with different maxBins and maxDepth, the decision tree models using incoming training data to train multiple models, and then the using evaluation data to evaluate the prediction results of each model. In this paper, the root means square error (RMSE) and the operation time are used to evaluate accuracy and computation cost respectively. According to the evaluation results, the model with high accuracy and relatively small cost operation is chosen as the optimal model. In the testing stage, the test data were passed in, and the optimal model was used for testing, and then the RMSE was calculated. Comparing the RMSE of the model between the evaluation stage and the test stage, if the difference is within acceptable range, it indicates that there isn't overfitting.

3.2 Feature Extraction

By long-term radio monitoring, the frequency channel occupancy of hours, days and months can be calculated, along with the maximum and average values of each time periods. Time, maximum level, average level and frequency channel occupancy constitute the basic data of frequency channel occupancy prediction. In each record of the basic data, frequency channel occupancy is the prediction target. The maximum level, average level and frequency channel occupancy are calculated at the same time, which cannot be used as the features. Time is the scale of frequency channel occupancy statistics, which can be used as a feature. However, using time as the only feature cannot fully show its relationship with the prediction target.

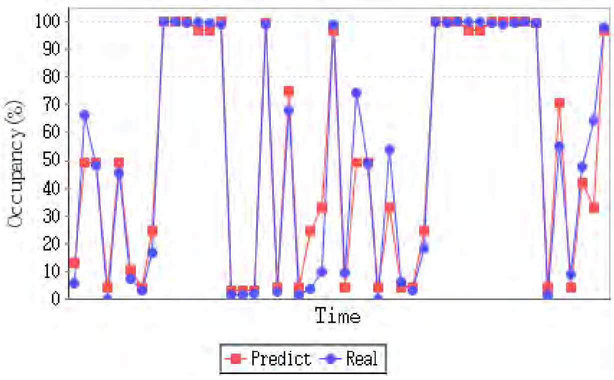
Because the emission time of some kinds of radio services has a certain correlation with which day of the week, and some other kinds with whether it's workday or not, the week sequence and whether holidays are two features derived from the time feature. In general, the frequency channel occupancy of a current moment has a strong relationship with it in the past m days. Therefore, this paper considers the frequency channel occupancy in the past m days as a feature, as well as its mean value, root mean square value and variance. In addition, the maximum level value and average level value calculated in the past n days are also taken as the features. Based on the hourly frequency channel occupancy data, the extracted feature and predicted target are shown in table 1.

Table 1. the extracted feature and predicted target of hourly frequency channel occupancy

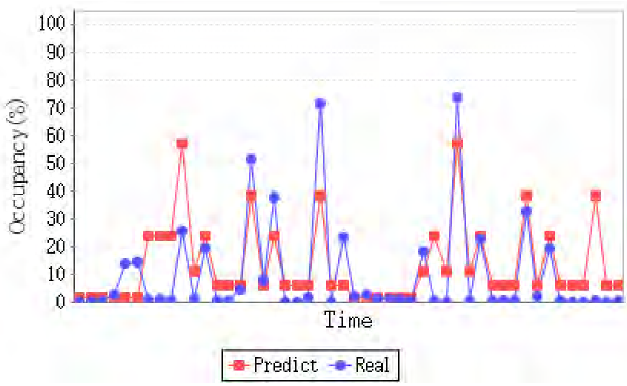
Feature	Format	Example
Time	year,month,day,hour	2018,6,24,17
Week sequence	which day of the week(0-6)	0
Whether it is holiday	True/false(1/0)	1
Frequency channel occupancy of the past m Days	frequency occupancy sequence(0-100%)	96.32%,98.25%,...
The calculated value of frequency channel occupancy of the past m day	mean value, root mean square value and variance	72.9,81.29,35.98
The maximum level value of the past n days	the maximum value(dbμv)	43.6,45.1,...
The average level value of the past n days	the average value(dbμv)	7.58,10.39,...
Predicted target label	frequency channel occupancy in current time(0-100%)	83.52%

4. Prediction of Radio Frequency Channel Occupancy

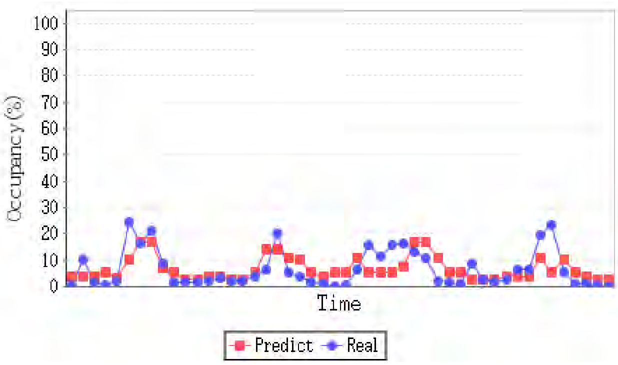
Some parts of the spectrum band on AM broadcast, amateur, maritime mobile and fixed services are chosen in this paper. Base on continuous monitoring frequency channel occupancy of up to 3 months statistical data, frequency channel occupancy (in last 7 days), the calculated mean value, variance, root mean square value, the maximum and average level value are chosen as features. On hadoop big data platform, Spark MLlib [7] is used to realize the tuning of decision tree model parameters to estimate the generated optimal model to predict the frequency channel occupancy of the next 24 hours. Fig. 3 shows the comparison between the predicted results and the actual results of the above two consecutive days of frequency channel occupancy.



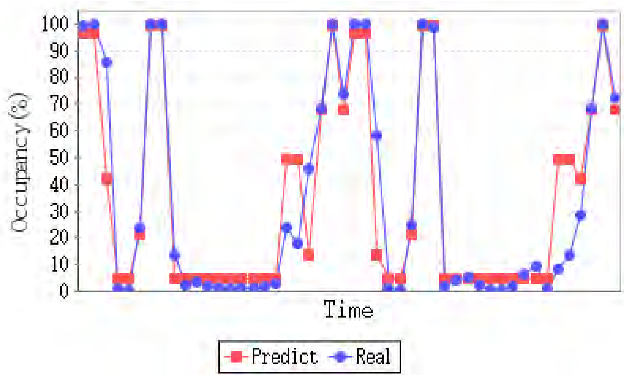
broadcast service (maxBins:10 maxDepth:5)



maritime mobile service (maxBins:50 maxDepth:3)



amateur service (maxBins:10 maxDepth:5)



fixed service (maxBins:100 maxDepth:5)

Fig.3 comparison chart of predicted frequency occupancy and actual values in different services

As figure 3 shows, in feature extraction of different historical frequency channel occupancy data, the maxBins and maxDepth parameters in training the optimal decision tree model are not necessarily the same. So, when predicting frequency occupancy, uniform parameters should not be used to train model, that is, parameter tuning is required in every single training of the optimal decision tree model. The test results also indicate that the predicted frequency occupancy of the four types of services are in good agreement with the actual test values and can reflect the its changing trend, which is of great help to master the efficiency of future frequency usage. In addition, when the predicted frequency occupancy greatly differs with actual monitoring values, there may be some new signals or interference, which is conducive to the estimation of abnormal frequency occurrence and has a certain practical significance.

To test the accuracy of frequency channel occupancy prediction by decision tree method, we select multiple frequencies in each service for seven days continuously. To compare the prediction results and the actual monitoring results, of which the statistical average RMSE, the maximum and minimum values are shown in table 2.

Table 2. the RMSE between the predicted frequency occupancy and the actual values

Type of radio service	average RMSE	maximum RMSE	minimum RMSE
AM broadcast service	21.76	30.98	10.56
maritime mobile service	18.22	25.71	11.70
amateur service	7.87	9.72	6.28
fixed service	11.71	15.02	9.71

It can be seen from table 2 that, due to different frequency occupancy of different radio services, especially differences in emission rules, the RMSE of predicted value and real value is somewhat different. On the whole, the prediction results have high accuracy as well as good practical value.

5. Summary

In this paper, big data technology is invoked in radio spectrum management. The classic decision tree in machine learning is used as the prediction algorithm to realize the prediction of frequency channel occupancy which is most commonly used in spectrum management. Decision tree algorithm is of great value for the efficient utilization of spectrum resources and also helpful for radio spectrum interference analysis in radio monitoring. The frequency occupancy prediction results of several different services show that the prediction is highly accurate. The feature extraction method, prediction method and decision tree parameter optimization process proposed here have has a certain scientific significance. They can be directly applied as practical technologies in radio management, which is also an exploration for intelligent radio management.

Acknowledgments

This work is supported by the National Science and Technology Major Project under Grants No. 2015ZX03002008.

References

[1]. Hongshun Zhou. Handbook of spectrum monitoring. People's posts and telecommunications press. 2006, p.1-3.

[2]. Information on: www. itu.int.

[3]. XingGong. Research and Implementation of Modulation Recognition Technology of Communication Signals, Master's thesis, Chongqing University of Posts and Telecommunications, China, 2017, p.62-63.

[4]. Xingyong Hong, Wenjing Li. Radio arrival bearing measurement algorithm based on decision tree. Electronics World,2016 No.9, p.80-81.

[5]. Da1i Li, Hangsheng Zhao, HongWei Liu, et.al. A policy reasoner Based on decision tree for processing spectrum management. National academic conference on radio applications and management 2010. Suzhou China,2010, p.40-45.

[6]. Dagui Lin. Hadoop and Spark big data massive analysis and machine learning. Tsinghua university press. 2017, p.8-10.

[7]. Sandy Ryza, Uri Laserson, Sean Owen, et.al. Advanced Analytics with Spark. People's posts and telecommunications press. 2015, p.4-10.