2018 18th IEEE International Conference on Communication Technology

A Method Based on Frequent Pattern Mining to Predict Spectral Availability of HF

Chujie Wu, Yunpeng Cheng, Yuping Gong, Guoru Ding, Ling Yu, Zhe Zhang
College of Communications Engineering
Army Engineering University of PLA
Nanjing, Jiangsu, China
e-mail: chujie128@163.com, chengyp2000@vip.sina.com, gyp78@sina.com, dr.guoru.ding@ieee.org, yl2811284434@126.com, zhangzhe850619@sina.com

***Abstract*—The HF radio communication has long been a big problem in channel selection since the spectrum environment is dynamic. To verify the feasibility of detecting idle channels by spectrum prediction, the data in this paper are based on real-world measurements collected by USRP in different time periods. The received signal power is converted to continuous sequences through a new channel state model reflecting spectrum availability. We then develop a prediction algorithm using simplified frequent pattern mining which can predict channel availability based on past channel states with considerable accuracy. The experimental results show that the measured data are more fluctuant in the afternoon which increase the predicted difficulty, nevertheless, the proposed algorithm is superior to neural network and Markov model in this situation, and the larger samples the better prediction performance.**

Keywords-HF; automatic link establishment; spectrum prediction; channel availability; frequent pattern mining.

I. INTRODUCTION

The High Frequency (HF) radio communication has been the earliest solution for wireless communication [1] with the characteristics of long-distance transmission and wide coverage [2]. Automatic Link Establishment (ALE) is the worldwide standard for HF communication, HF radio system can select the optimal channel automatically through ALE mechanism in the dynamic spectrum environment.

Since the late 1970s, there had been data transmission methods for HF radio like selecting channel or tuning frequency manually, when the ALE did appear. In the mid-1980s, a military standard for ALE called 2G-ALE [3] with the function of Link Quality Analysis (LQA) was accepted and it's still widely used today. In the mid-1990s, to overcome the shortcomings of 2G-ALE, 3G-ALE was introduced as newest protocol for faster and more reliable link establishment [4]. Both 2G-ALE and 3G-ALE have a limited number of channels (typically 10 to 20 frequencies) [5] for the link establishment. These frequencies in so called ALE groups need to be defined manually through channel detection and can only be taken into account under static propagation condition [6]. In order to improve the usability of HF radio systems, the 4G-ALE is being investigated by corporations and military. The related studies about 4G-ALE mainly involve cognitive radio and wideband transmission [7].

In cognitive ALE, for the purpose of exploiting the spectrum holes, the radio is aware of its environment and able to adjust the operation parameters [8]. The work in [9] provides an overview of the techniques in cognitive radio and considers the challenges of applying these techniques to the HF radio system. One of these challenges is to create system prediction model for estimating noise and interference on the frequency spectrum. The ALE time is reduced greatly by prediction, since the number of sensed channels are narrowed down. Besides, once the prediction accuracy reaching a very high level, the system can skip the sensing module to establish the link directly. The work in [10] introduces the ITS HF Propagation, an international HF propagation link simulation model which can predict the Maximum Usable Frequency (MUF), median of field strength, Signal to Nosie Ratio (SNR) between two points, but it is long-term prediction applicable for months. Haris [11] developed a Neural Network model to predict the 24-hour variation of occupancy within each HF Broadcast allocation. But the data set used for model was measured for several years, and the optimum number of hidden units was determined by trial and error. The work in [12] developed a 2D frequent pattern mining algorithm that can predict channel availability based on past observations in 20MHz to 3GHz, but not quite related works to HF band.

The main contributions of this paper are summarized as follows:

- Propose a new definition of channel availability from the perspective of interference degree in a period of time, and divide the channel states into unavailable, medium and best.
- Develop a simplified frequent pattern mining (FPM) algorithm with the key idea of finding the most frequent subsequences in the data set through pattern recognition and exploring their association rules to predict future channel availability.
- Present real-world experiments of HF band to demonstrate the effectiveness of the developed algorithm. The main insights include spectrum varies from time to time, but the proposed algorithm predict better in some cases.

The reminder of the paper is organized as follows: Section II analysis the spectral characteristics of measurements and presents a new channel state model. In Section III, a simplified FPM algorithm to predict spectrum states is developed. In Section IV, experimental results and discussion under different situations are compared. In Section V, we draw conclusions about this paper.

II. Data Collection and Channel State Modeling

A. Data Collection

The results presented in this paper are based on measurement data of HF collected on the same day from 10:00 to12:00 (morning) and 16:00 to 18:00 (afternoon) in Nanjing (E118° 50', N32° 01'). The equipment we used is a Universal Software Radio Peripheral (USRP), which is connected to the whip antenna via coaxial cable (Fig. 1). The computer controls the USRP through LAN, and the received signal power and sampling time of each channel are recorded in the PC in the form of Excel.

The measurement resolution is 0.1MHz, there are 270 channels within the frequency range of 3~30MHz. In order to better reflect the energy level changing of frequencies, we set the time slot to 0.4s, that is, the variation of the spectrum is recorded every 0.4s. As a result, there are 3.6×10^4 (240mins/0.4s) time slots and $3.6 \times 10^4 \times 270$ readings in the data set.

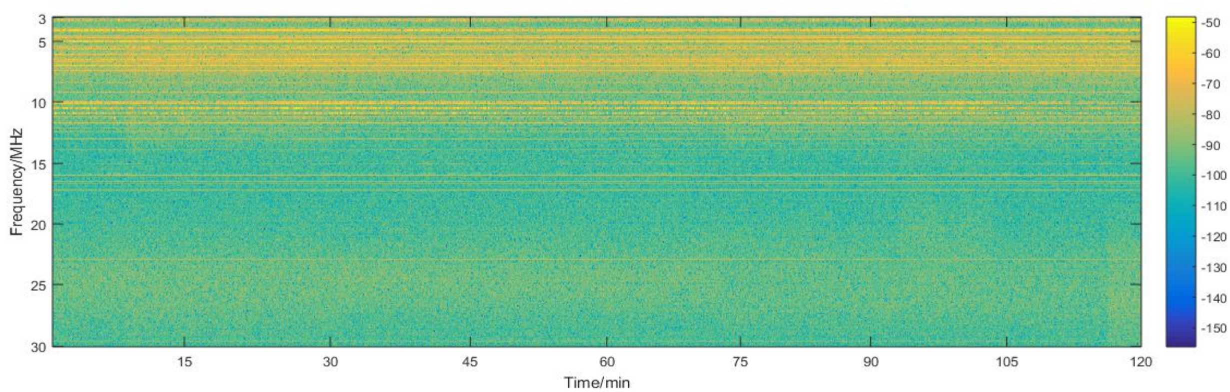
The HF spectrum is hardly same all the day, that's because the ionosphere layer isn't stable. Fig. 2 shows the spectral power evolution maps. The interference is worse in the afternoon, and the number of frequencies of good quality have been decreased.

Figure 1. System architecture of the measure equipments

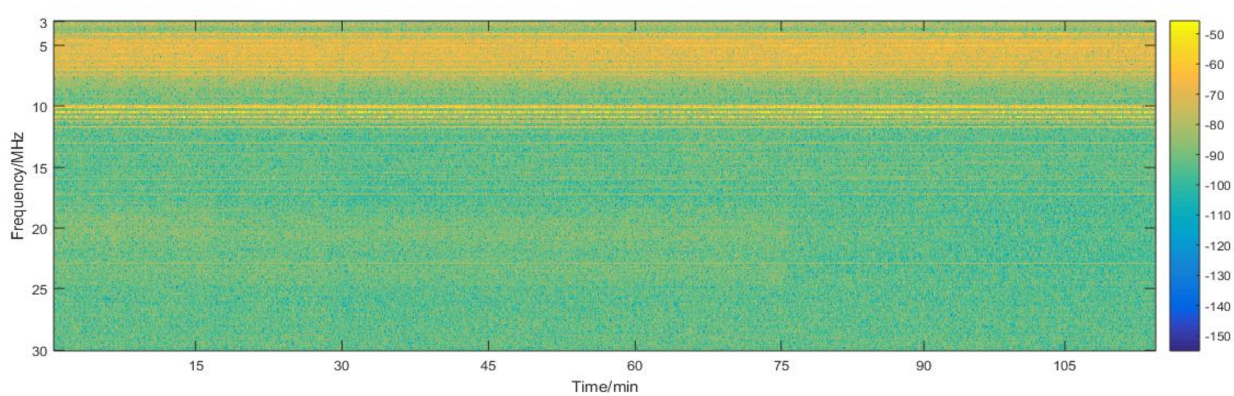
Furthermore, we denote the received signal power as P_r, the correlation of P_r between adjacent minutes is given by

$$\gamma = corr(P_r^t, P_r^{t+1}) \tag{1}$$

where $corr$ refers to the correlation coefficient. We plot γ in Fig. 3, it can be seen from the picture that the measured data from 16:00 to 18:00 are more fluctuant due to the effect of sunset, and we presume that P_r isn't only related to the previous moment.



(a) 10:00-12:00



(b) 16:00-18:00

Figure 2. Spectral power evolution maps

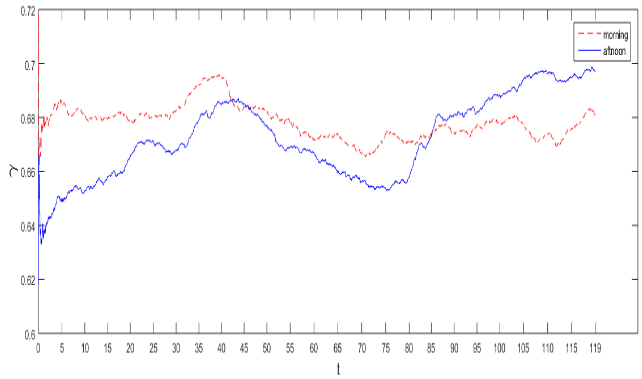


Figure 3. Correlation coefficient between adjacent minutes

B. *Channel State Modeling*

The existing modeling approach of channel state is to compare the received signal power to a threshold, with 1 denoting a channel being busy when the power exceeds the threshold, and 0 denoting the opposite [13]. However, we pay more attention to the interference over a period of time. For instance, if there are more signals with high power during this period, indicating that the noise is serious at the moment, if the high power signals are few, the channels are suitable for access. Therefore, we quantify the signals from time and power in reference to the thresholds in [14] to divide the channels into three states including unavailable, medium and best.

The channel occupancy probabilities of three intervals ($[+\infty,-77dB]$, $[-77dB,-97dB]$, $[-97dB,-\infty]$) in the unit of d seconds are denoted by $P_{unavailable}$ P_{medium} and P_{best}, which can be obtained through calculating the time ratio in each interval,

$$\begin{cases} P_{unavailable} = \dfrac{T_{(\mathrm{Pr}\geq -77dB)}}{d} \\ p_{medium} = \dfrac{T_{(-97dB > \mathrm{Pr} > -77dB)}}{d} \\ p_{best} = \dfrac{T_{(\mathrm{Pr}\leq -97dB)}}{d} \end{cases} \tag{2}$$

To illustrate the channel states, we define q_n^t as a function of channel n at moment t by comparing the above channel occupancy probabilities,

$$q_n^t = \begin{cases} 3, & P_{unavalable} \geq 0.5 \quad Unavailable \\ 2, & P_{medium} > P_{best} \quad Medium \\ 1, & P_{medium} < P_{best} \quad Best \end{cases} \tag{3}$$

$$t_{\max} = 240\min \times \frac{60\sec}{d}, n \in N = \{1,2,3...270\}, t \in T = \{1,2,3...t_{\max}\}$$

We can presume larger q_n^t indicates smaller channel's availability.

On the basis of measured time, denoting the $N \times t$ channel state matrix as Q_N^t,

$$Q_N^t = \begin{pmatrix} q_1^1 & q_1^2 & \cdots & q_1^t \\ q_2^1 & q_2^2 & \cdots & q_2^t \\ \vdots & \vdots & & \vdots \\ q_N^1 & q_N^2 & \cdots & q_N^t \end{pmatrix} \tag{4}$$

We convert the spectral power to continuous sequences through the above modeling method, the data volume is reduced and getting the effective information about spectral availability. It's better for channel selection and channel access in HF. Furthermore, the collected data is modeled into channel states to conduct the prediction presented in the next section

III. PREDICTION USING FREQUENT PATTERN MINING

A. *Prediciton by FPM*

The time correlation presented in the above section suggests that we can predict future spectral condition base on measurements made in the past. There are different prediction models such as Neural Network [14], Markov [15], Support Vector Machine (SVM) [16], etc. However, the Neural Networks have a high demand to the number of samples and the choice of network architecture affects the outputs directly, the Markov takes great time to calculate the transition probability and the SVM has a high complexity. To overcome these advantages, we develop a new approach using FPM which can we predict the Q_N^{t+1} on the basis of Q_N^t.

Frequent patterns are item sets, subsequences, or substructures that appear in a data set with frequency no less than a user-specified threshold [17]. Frequent pattern mining was first proposed by Agrawal et al. in 1993 for market basket analysis. It analyses customers' buying habits by finding associations between different items in their shopping baskets [18]. The prediction of channel state through FPM is to find the most frequent submatrices in the data set and explore their relevance. For example, suppose we know the states of channel *n* before moment *t*, which can be written into a matrix $Q_n^{t-1} = [q_n^1 \ q_n^2 \ ... \ q_n^{t-1}]$ define a submatrix as a pattern if it appears no less than 30 times in Q_n^{t-1}. If we find the pattern [3 2 1] appears 50 times, while another pattern [3 2 1 2] appears 49 times, we can predict the probability $P(q_n^t = 2) = \dfrac{49}{50}$ when $[q_n^{t-3} \ q_n^{t-2} \ q_n^{t-1}] = [3 \ 2 \ 1]$.

B. *Algorithm Design*

There are two steps to predict channel states by FPM: one is to find frequent patterns, the other is to mine the association rules [19]. The existing FPM algorithms such as Apriori and FP-growth [20] are suitable for very huge number of items where the association rules aren't very obvious. However, the frequent patterns in this paper are sequences come from HF which are more regular and easier to mine the relevance. To reduce the complexity, we develop a simplified FPM prediction algorithm. The key idea is to find the required pattern and all of its sub-patterns in the data set through the pattern recognition process, the sub-pattern appears most frequently is connected with future channel state.

The pseudo-code of the proposed algorithm is given in Algorithm 1, F_L is the required frequent pattern with the subsequence length of L. We say f_{L+1} is F_L's sub-pattern if the front L is the same with F_L but the $L+1$ is different such as $F_L=[1\ \ 2], f_{L+1}=[1\ \ 2\ \ 1]$. According to the classification of channel states in section Ⅱ, every F_L having at most three f_{L+1} written as $f_{L+1}^i=[F_L, i]\ (i=1,2,3)$. We start by finding the frequent pattern F_L from the last L states in Q_n^t and search its sub-patterns f_{L+1}^i. Then we count each sub-pattern's appearance number denoted as s_i in the Q_n^t. If the f_{L+1}^i with the maximal s_i which is greater than the minimum support s_m, the prediction result is the i. We can also put Q_N^{t+1} to the historical data to predict multiple slots.

Algorithm 1: simplified FPM prediction algorithm
input: Q_N^t for n=1:N $Q_n^t=[q_n^1\quad q_n^2\quad \cdots\ q_n^{t-1}\quad q_n^t]$ $F_L=\left[q_n^{t-L+1}\ \ldots\ q_n^{t-1}\ \ q_n^t\right]$ $f_{L+1}^i=[F_L\ ,\ i]\quad (i=1,2,3)$ $s_i=appearance\_number\,[f_{L+1}^i, Q_n^t]$ if $\max(s_i)>s_m$ then $q_n^{t+1}=i\,\vert_{\max(s_i)}$ end end $Q_N^{t+1}=\left[Q_N^t, q_N^{t+1}\right]$ **output:** Q_N^{t+1}

IV. Experimental Results and Discussion

To test the algorithm, the collected measurement data are modeled into channel states, which are split into two parts, one as a training set to run the algorithm, while the other as the testing set to compare the prediction performance. In a fixed number of data, f_{L+1}^i with greater L have fewer s_i. In order to find the maximal s_i and optimal i, we set L=2, d=30s.

A. Different Methods' Performance

The spectral characteristics presented in section Ⅱ show the data are more stable in the morning. Therefore, based on the channel sates from 10:00 to 12:00, we compare the prediction accuracy of BP Neural Network model, Markov model and the proposed FPM algorithm. The training set is 90 minutes, to predict the next 5 to 30 minutes. The comparison results of three methods are shown in Fig. 4. It's seen that the FPM algorithm and Markov model have small difference in accuracy, while the Neural Network model isn't suitable for short-term prediction on HF due to the limited data samples.

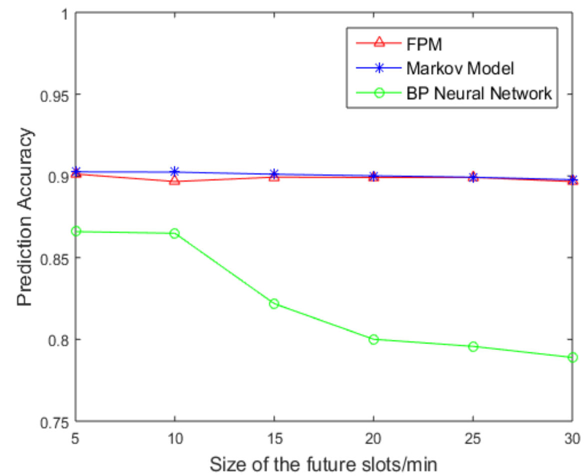
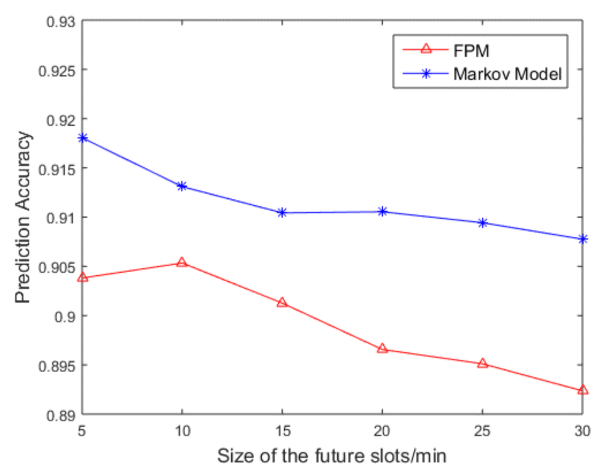


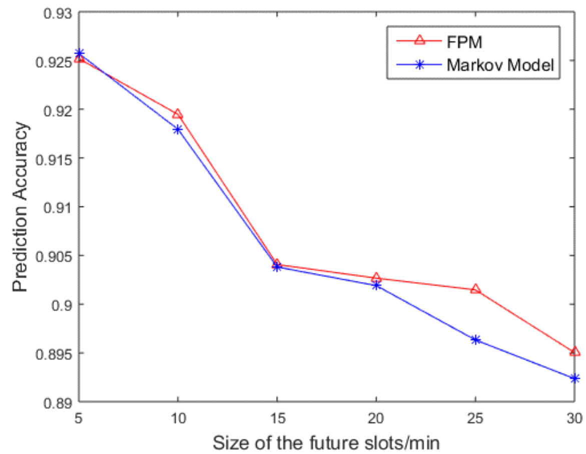
Figure 4. Comparison of prediction performance of different methods

B. Different Periods' Performance

To compare the Markov model and FPM algorithm further, we run the two methods under same training set but different measured time. From Fig. 5 we observe that, the prediction accuracy decreases with the increase of time span. Furthermore, compared to the morning, the accuracy of both methods decreases more significantly in the afternoon, that's because the interference is severer and the signals are more variable. Furthermore, the FPM method is slightly better than Markov under more dynamic spectrum condition.



(a) 10:00-12:00



(b) 16:00-18:00

Figure 5. Comparison of prediction performance in different measured periods

C. *Different Samples' Performance*

To analyze the impact of sample size on prediction performance, we select the training set of 30minutes, 60minutes and 90minutes respectively in the morning to predict the channel states in the next 30 minutes by proposed FPM algorithm. The results in Fig. 6 suggest that as the time increase, the curve is more stable when getting more training samples. Because more training set capture the changing spectral characteristics better, and can stabilize the predictive accuracy.

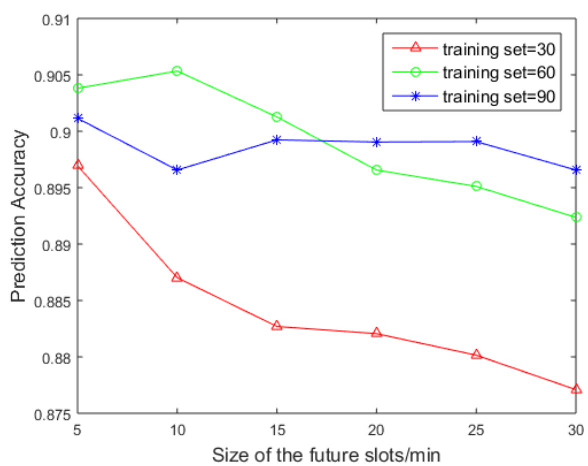


Figure 6. Comparison of prediction performance under different training set

V. CONCLUSION

In this paper, we carried out spectrum measurement of HF at different time periods based on the USRP equipment. Using these data we conduct analysis of spectrum characteristics and convert the data into continuous channel states by a new spectrum availability model. Moreover, we develop a simplified FPM prediction algorithm, which can find the needed frequent patterns through the pattern recognition process and extracting the association rules. The simulation results prove that the proposed FPM algorithm is superior to other prediction models in HF especially when the spectrum varies greatly, and the larger samples the better prediction performance.

REFERENCES

[1] J. Wang, "Research and Development of HF Digital Communications (in Chinese)," *Beijing: Science Press*, 2013.

[2] A. Shukla, N. Jackson-Booth, P. Arthur, "Cognitive Radios' and their relevance to HF radio systems," in *Proc. IET International Conference on Ionospheric Radio Systems and Techniques*, 2012, pp. 1-6.

[3] M. Baker, W. Beamish, M. Turner, "The use of MIL-STD-188-141A in HF data networks," in *Proc. IEEE Military Communications Conference (MILCOM 89)*, 1989, pp. 75-79.

[4] E.-E. Johnson, "Third-generation technologies for HF radio networking," in *Proc. IEEE Military Communications Conference (MILCOM 98)*, 1998, pp. 386-390.

[5] T. Vanninen, T. Linden, M. Raustia and H. Saarnisaari, "Cognitive HF — New perspectives to use the high frequency band," in *Proc. 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, 2014, Oulu, pp. 108-113.

[6] L. Economou, H. Haralambous, C. Pantjiaros, P. Green, G. Gott, P. Laycock, M. Bröms, and S. Boberg, "Models of HF spectral occupancy over a sunspot cycle," *IEE Proceedings—Communications*, vol. 152, no. 6, pp. 980–988, 2005.

[7] Z. Qin, J. Wang, J. Chen, G. Ding, "Link quality analysis based channel selection in high-frequency asynchronous automatic link establishment: a matrix completion approach," *IEEE Systems Journal*, vol. 12, no. 2, pp. 1957-1968, June 2018.

[8] A. Shahid, S. Ahmad, A. Akram and S. A. Khan, "Cognitive ALE for HF Radios," *in Proc. Second International Conference on Computer Engineering and Applications*, 2010, Bali Island, pp. 28-33.

[9] E. Koski and W. N. Furman, "Applying cognitive radio concepts to HF communication," in *Proc. IET International Conference on Ionospheric Radio Systems and Techniques*, Apr. 2009, pp. 1–5.

[10] J. Zhao, N. Zhao, "An overview of the international HF transmission link simulation model and its application analysis(in Chinese)," *China Radio*, vol.1, pp.47-49, 2017

[11] H. Haralambous and H. Papadopoulos, "24-Hour Neural Network Congestion Models for High-Frequency Broadcast Users," *IEEE Transactions on Broadcasting*, vol. 55, no. 1, pp. 145-154, March 2009.

[12] S. Yin, D. Chen, Q. Zhang, M. Liu and S. Li, "Mining Spectrum Usage Data: A Large-Scale Spectrum Measurement Study," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 1033-1046, June 2012.

[13] L. Yu, J. Chen, G. Ding and Z. Qin, "Fast automatic link establishment: A new metric and the value of spectrum prediction," in *Proc. 8th International Conference on Wireless Communications & Signal Processing (WCSP)*, 2016,Yangzhou, China, pp. 1-6.

[14] H. Haralambous and H. Papadopoulos, "24 HF Spectral Occupancy Characteristics and Neural Network Modeling over Northern Europe," *IEEE Transactions on Electromagnetic Compatibility*, vol. 59, no.1, pp. 1817-1825, 2017.

[15] C. Ghosh, C. Cordeiro, D. P. Agrawal, and M. B. Rao, "Markov chain existence and hidden Markov models in spectrum sensing," in *Proc. IEEE International Conference on Pervasive Computing and Communications*, 2009, Galveston, TX, pp. 1-6.

[16] S. Ni and S. Shen, "Frequency spectrum access mechanism of cognitive radio based on spectrum prediction," in *Proc. IET*

International Conference on Communication Technology and Application (ICCTA), 14-16 Oct. 2011, Beijing, China, pp. 556–560.

[17] J. Han, H. Cheng, D. Xin, and X. Yan, "Frequent Pattern Mining: Current Status and Future Directions," *Data Mining and Knowledge Discovery,* vol. 15, no. 1, pp. 55-86, 2007.

[18] X. Zhu, H. Deng and Z. Chen, "A Brief Review on Frequent Pattern Mining," in *Proc. 3rd International Workshop on Intelligent Systems and Applications*, 2011,Wuhan, China, pp. 1-4.

[19] G. Cong, K.-L. Tan, A. Tung, and F. Pan, "Mining Frequent Closed Patterns in Microarray Data," in *Proc. Fourth IEEE International Conference on Data Mining (ICDM '04)*, Nov. 2004, pp. 363-366.

[20] P.-N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, pp. 208-210. Addison Wesley, 2005.