Comparison of the Long-Term forecasting method of RSSI by Machine Learning

Tatsuya Nagao\(^{(a)}\), Takahiro Hayashi\(^{1}\), and Yoshiaki Amano\(^{1}\)

\(^{1}\) KDDI Research, Inc., Fujimino-shi, Saitama, 356-8502, Japan
\(a)\) tt-nagao@kddi-research.jp

Abstract: In order to improve the efficiency of spectrum use, systems that share spectrum while avoiding interference between different systems are being investigated. In the millimeter-wave band, which is expected to be utilized in the future, the received power fluctuates due to quasi-static obstructions such as people and vehicles, but such temporal variations have not been taken into account in conventional methods. In this paper, we use a variety of machine learning algorithms for comparative evaluation to forecast the temporal fluctuations of radio propagation due to changes in the number of people and vehicles in order to achieve more dynamic spectrum access.

Keywords: dynamic spectrum access, machine learning, time-series analysis

Classification: Wireless communication technologies

References

[1] M. Matinmikko et al., “Spectrum sharing using licensed shared access: the concept and its workflow for LTE-advanced networks,” in IEEE Wireless Communications, vol. 21, no. 2, pp. 72-79, April 2014.
[2] M. M. Sohul, M. Yao, T. Yang and J. H. Reed, “Spectrum access system for the citizen broadband radio service,” in IEEE Communications Magazine, vol. 53, no. 7, pp. 18-25, July 2015.
[3] Y. Chen and H. S. Oh, “A survey of measurement-based spectrum occupancy modeling for cognitive radios,” IEEE Commun. Surv. Tutorials, vol. 18, no. 1, pp. 848–859, 2016.
[4] F. Pedregosa et al., “Scikit-learn: Machine learning in Python”, J. Mach. Learn. Res., vol. 12, pp. 2825-2830, Oct. 2011.
[5] T. Chen et al., ”XGBoost: A scalable tree boosting system”, Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 785-794, 2016.
[6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997.
[7] ITU, “Guidelines for evaluation of radio interface technologies for IMT-2020,” Rep. ITU-R M.2412-0, 2017.
1 Introduction

With the rapid increase in mobile data traffic, the allocation of spectrum resources has become an urgent issue. However, since many frequency bands are already used by various radio systems, systems that share spectrum between different systems (dynamic spectrum access: DSA) are being considered. In DSA systems, it is important to understand the coverage of the primary systems. However, in the millimeter-wave band, which is expected to be utilized in the future, the coverage is more sensitive to blockers such as pedestrians and vehicles, and the coverage fluctuates as the number of blockers changes through time. Conventional DSA systems, such as Licensed Shared Access (LSA) in Europe [1] and Spectrum Shared Access (SAS) of the Citizens Broadband Radio Service (CBRS) in the United States [2], do not take into account this kind of time fluctuation and are based on a conservative approach that assumes worst-case scenarios. In order to further increase frequency utilization efficiency through more dynamic sharing, it is important to forecast long-term changes in coverage that take into account the impact of blockers.

Recently, machine learning has been applied in the field of time series prediction, but it is necessary to clarify what kind of prediction model is suitable for the fluctuations of radio propagation due to quasi-static blockages such as those described above. In this paper, we conduct a comparative evaluation of the long-term forecasting of the received signal strength indicator (RSSI), which varies on an hourly basis, using various machine learning algorithms.

2 RSSI forecasting by Machine Learning

Time series forecasting is used in a wide range of fields, including natural sciences, industry, business, and economics. A conventional method of time series analysis is the statistical model represented by the Auto Regressive Integrated Moving Average (ARIMA) [3]. Since time series forecasting is also a regression problem based on past data, machine learning can be applied to it. In other words, we can use past time series data as a feature vector to obtain future predictions by regression. In this paper, we have used the following five machine learning algorithms, which are commonly used for regression tasks.

- Ridge Regression [4]
- Support Vector Regression [4]
- Random Forest [4]
- XGBoost [5]
- Long Short-Term Memory (LSTM) [6]

We used standard parameters for each algorithm. The summary and main parameters of each algorithm are described below. Ridge Regression is a linear regression model, and it suppresses over-fitting by regularization using the L2 norm. We set the regularization strength $\alpha = 1.0$. Support Vector Regression enables nonlinear regression through a technique called the kernel trick, which
extends the feature space to higher dimensions. We used Radial Basis Function (RBF) as the kernel, the regularization factor \( C = 1.0 \) and the insensitive loss function \( \varepsilon = 0.1 \). Random Forest and XGBoost are called ensemble learning methods, which integrate multiple weak learners to build a single prediction model. Random Forest performs weak learner training independently (Bagging: Bootstrap Aggregating), while XGBoost uses the results of the previous weak learner and trains the next weak learner sequentially (Boosting). For both algorithms, we set \( n_{\text{estimators}} = 100 \), which represents the number of weak learners, and \( \text{max_depth} = \text{"None"} \), which represents the depth of the regression tree that is a weak learner (meaning no restrictions). LSTM is a type of Recurrent Neural Network (RNN) and has functions to control weights while retaining information about past data - it is often used in natural language processing. Although more layers allow for more complex expressions, network structure and parameter tuning is difficult and the computational cost is large. We configured the LSTM network with 5 layers with 12 units, and used “tanh” (hyperbolic tangent function) as the activation function.

3 Evaluation
We generated the time series of RSSI data, which fluctuates due to obstructions such as people and vehicles by simulation, and evaluated the forecasting accuracy by various machine learning algorithms.

3.1 Simulation
We conducted the simulation in an urban area and the frequency was 26 GHz. It has 1 transmitting station (0 dBm transmission power, omni-antenna) and 4 receiving points. The location of the receiving points is not changed on the assumption that fixed-point measurements can be made using such as spectrum sensors. The average of the forecasting accuracy for each of the 4 points was used to evaluate. The procedure and an example result are shown in Fig.1. First, we obtained path information for each receiving point by performing a ray tracing simulation considering static buildings. Next, we generated the hourly average number of blockers for each receiving point from the actual LTE system log data for 30 days. This log data includes the times and locations of the UEs connected to the LTE system. Next, the hourly average of the number of blockers within 25m was generated for each receiving point based on the pattern of variation in the LTE system logs. Then, the number of blockers was determined according to the Poisson distribution, and the attenuation was calculated stochastically for each path according to the ITU-R M. 2412 Blockage Model I [7]. We calculated this 60 times and took the median value as the typical value for a timestamp. By the above procedure, the hourly and 30-day RSSI time series data were generated. As this paper provides forecasting of fluctuations, please note that the absolute value of the RSSI itself has no direct impact on the results of the evaluation.
3.2 Results
For evaluating the forecasting accuracy, we used the Root Mean Squared Error (RMSE). The RMSE of the forecasted value $\hat{y}_i (i = 1, \cdots, N)$ for the observed value $y_i$ is calculated by the following equation.

$$ \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} $$

Of the 30-day data, the first 7 days were used for training and the remaining 23 days were used for forecasting and evaluating. Also, to assess the long-term forecasting, which is the purpose of this paper, the number of forecasting steps was increased from 1 to 48. This means that we are making forecasts from an hour ahead to two whole days ahead. Here, the LSTM network structure is composed of 5 layers of 12 units.

The results of each machine learning algorithm are shown in Fig. 2. For comparison, the results of the statistical model ARIMA are also included. From the figure, we can see that machine learning is effective in long-term forecasting. Also, among the machine learning methods, Random Forest and XGBoost, which are ensemble learning methods, are highly accurate and maintain stable accuracy for long-term forecasting.
4 Discussion

In this paper, we assume that blockers such as people and vehicles cause fluctuations in the received power. And the number of blockers changes periodically according to people's activity patterns. Therefore, the fluctuations of the received power are also periodic. The autocorrelation of the simulated RSSI time series data is maximal at lag = 24, meaning that it generally fluctuates over a 24-hour period.

Basically, the accuracy declines as the forecasting target becomes further into the future. But as can be seen from Fig. 2, RMSE increases until the number of forecasting steps reaches 6, and has not changed much since then. This can be explained by the fact that the minimum autocorrelation is at lag=6. Since the autocorrelation becomes larger after the 6th step, it is considered that the past periodic pattern can be learned and reflected in the forecasting value.

Fig. 3 shows the feature importance of the Random Forest model, which had the best RMSE. Feature importance indicates how much the input features contribute to the improvement in accuracy. The vertical axis shows how many steps ago the input data was, and the horizontal axis shows the number of forecasting steps. From the figure, there is a clear periodicity between the feature importance of the number of lags of the input data and the number of forecasting steps. The interval between the lag of the highly important input data and the number of forecasting steps is around 24, which corresponds to the period time of the time series data described above. Therefore, it was found that Random Forest can accurately learn the periodicity of time series data.
5 Conclusion
In this paper, we use various machine learning algorithms to make long-term forecasts of the fluctuations of the received power, considering the blockage attenuations caused by people and vehicles. A comparison with ARIMA, a leading statistical model, has shown the effectiveness of machine learning. In addition, we found that ensemble learning, such as Random Forest and XGBoost, can provide stable and accurate long-term forecasting. In this case, the fluctuations were periodical, and it is necessary to develop methods for forecasting and detecting non-periodic events that may occur in real environments. We also plan to conduct parameter tuning of each machine learning algorithm to improve forecasting accuracy.

Acknowledgments
This research is supported by the Ministry of Internal Affairs and Communications in Japan (JPJ000254).