Machine Learning for Performance Prediction in Mobile Cellular Networks

Janne Riihijärvi and Petri Mähönen
Institute for Networked Systems, RWTH Aachen University, Aachen, Germany

Abstract—In this paper, we discuss the application of machine learning techniques for *performance prediction* problems in wireless networks. These problems often involve using existing measurement data to predict network performance where direct measurements are not available. We explore the performance of existing machine learning algorithms for these problems and propose a simple taxonomy of main problem categories. As an example, we use an extensive real-world drive test data set to show that classical machine learning methods such as Gaussian process regression, exponential smoothing of time series, and random forests can yield excellent prediction results. Applying these methods to the management of wireless mobile networks has the potential to significantly reduce operational costs while simultaneously improving user experience. We also discuss key challenges for future work, especially with the focus on practical deployment of machine learning techniques for performance prediction in mobile wireless networks.

1556-603X/18©2018IEEE

I. Introduction

Automating the optimization and management of wireless mobile networks has the potential to significantly reduce the operational costs for network operators, as well as to improve the quality of user experience. Currently, much of network management involves human interaction ranging from conducting drive tests in order to evaluate the network coverage and performance to diagnosing customer complaints. We argue that machine learning techniques can assist in these tasks by reducing the need for drive tests, and helping to predict and diagnose network failures even before they noticeably degrade the quality of service of the network users.

In this paper we focus on *performance prediction* problems related to mobile network optimization and management. These generally involve predicting the network performance at locations or times from which no direct measurement data is available. We discuss the main categories of such prediction problems and demonstrate how they can be solved with state of the art machine learning methods using a real-world data set on measured mobile network performance as a source of examples. We also discuss extensively future research directions and challenges arising from this article that will need collaboration between the computational intelligence and wireless communications communities for their solution.

Despite the central role wireless networks play in today's society, real-world applications of machine learning for their operations have remained surprisingly sparse. Some examples of early work on performance prediction include [1]–[5], but these rely critically on the use of simulated data or small dedicated test networks, and have focused only on small network deployments compared to nation-wide networks mobile operators need to manage. Security related applications have been more thoroughly studied [6], [7], but these have very different nature from the problems we consider, usually extending considerably beyond the wireless network itself. For wireless sensor networks numerous machine learning applications have been suggested for characterizing the state of the network and the phenomena being sensed, but real-world deployments of such systems have again remained elusive [8]–[10].

The main research domain where machine learning has been applied to wireless networks has been the work to develop so-called *cognitive radios* and *cognitive wireless networks* [11]–[14], including suggestions of generic architectures on employing machine learning for cognitive networking [15]–[17]. Here the key concept that has emerged during the last decade is that of a *radio environment map* (REM) [18], [19] which can be thought of as a database where cognitive wireless network nodes store the information they have gathered of the environment, and which cognitive resource management solutions (based on machine learning and classical optimization approaches) can base decisions on. In these terms the present work can be seen as an exploration how to process and enrich the data stored in a REM. Applications of REMs for cellular networks are numerous (see [18]), including improved estimation of cell boundaries and handover possibilities between base stations, estimating interference relationships between different transmitters, as well as processing of possibly highly unreliable measurement data collected by mobile terminals to complement drive tests.

The rest of this paper is structured as follows. In Section II we introduce an example data set for our case studies, together with the types of prediction problems considered throughout the rest of the paper. In Sections III through V we then study how machine learning methods perform on these prediction problems using our example data set. In Section VI we outline challenges for future work, especially related to deploying machine learning predictors in practice for wireless network optimization and management. Finally, we conclude the paper in Section VII.

II. Case Studies and Prediction Problems

In this section we will briefly introduce the data set used for the rest of the paper as a source of case studies, as well as introduce the different prediction problems underlying them. We introduce our example data set early on, as we believe that the best approach is to explain problems with concrete case studies with real data, and thereby ground the subsequent discussion on machine learning methods to realistic applications in mobile wireless networks.

A. Example Data Set

We focus on the data set shown in Figure 1, which corresponds to a 10 km × 10 km subset within a large professional drive test campaign conducted in a major metropolitan area in the US. The full data set covered an area of approximately 120 km × 170 km and consisted of 4 million measurement locations in total. The gathered measurements are all for the performance characteristics of a 2 GHz downlink band for a major 4G cellular network operator. The variables included in the data set are geographical coordinates and time of the measurement, the received signal strength (RSS) measured at each location in dBm, the carrier to interference (C/I) ratio measured in dB, as well as the achieved user data rate (in bps) for a simple data download application. Based on these data we further enrich the data set with information on the velocity of the mobile drive test unit, as this is expected to play a role especially in the achieved data rate.

From Figure 1 we see that the three variables illustrated in the individual panels have very different behaviors both in terms of their *spatial distribution* as well as in terms of their relationships with each other. RSS varies relatively smoothly over the drive test region, whereas the C/I ratio has somewhat less smooth behavior. Nevertheless, the two are strongly correlated

Digital Object Identifier 10.1109/MCI.2017.2773824
Date of publication: 10 January 2018

Corresponding Author: Janne Riihijärvi (Email: jar@inets.rwth-aachen.de).

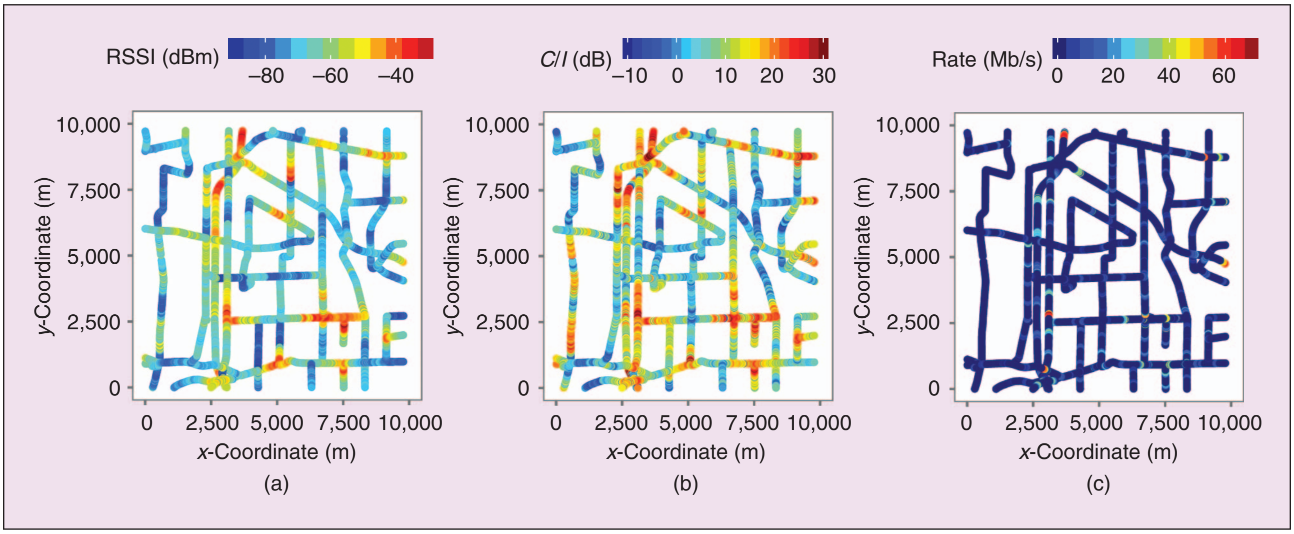

FIGURE 1 Example data set used throughout the paper as basis for case studies. (a) Received signal strength, (b) *C/I* ratio, (c) Data rate.

as can be expected. The achieved data rate, on the other hand, has very erratic behavior, with extreme variability over space and being only loosely correlated with the other two variables. Theoretically, the achievable data rate is, of course, determined by the C/I ratio, but in practice imperfections introduced by the various protocols layers as well as possible limitations in the backhaul capacity can play a dominant role, as evidenced by this example data set.

B. Types of Prediction Problems and Methods to Solve Them

We shall now introduce three categories of prediction problems, all of key relevance to the operations and management of commercial mobile wireless networks. This categorization is not meant to be exhaustive, although vast majority of problems we have encountered fall into one of these categories. We shall also briefly introduce selected machine learning methods whose effectiveness in solving these problems we will demonstrate in the rest of the paper. For a comprehensive introduction to machine learning, we refer the reader to [20] and references cited therein.

The objective of all of the methods considered here is to find a function that *predicts* the value of a dependent variable $y = f(x_1, \ldots, x_n)$ as a function f that depends on various *predictors* or *covariates* x_i. Usually, this is done by conducting a limited number of experiments that yield the value of y for *known* values of the covariates that are then used to *train* the model. In our case the variables shown in Figure 1 together with the location and velocity data form the collection of all the available covariates as well as potential choices of y.

1) Spatial Prediction Problems

We begin by considering a *spatial prediction* problem, where x_i are measured values of, for example, received power at selected n locations, and y ranges through locations at which no measurements are available. Such problems are especially important in extending results from drive tests into nearby areas not covered by the drive[1]. Particularly powerful prediction method is obtained by applying *Gaussian process regression* [21], which is closely related to *kriging* developed in the spatial statistics community [22] and applied to wireless networks in the REM context in, e.g., [23], [24]. While essentially any regression method can be used for spatial prediction, methods inspired by interpolation approaches tend to work in practice the best, and from these Gaussian process regression excels by enabling essentially all parameters involved to be estimated directly from the data.

Gaussian process regression treats the measured values as samples $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ from a realization of a random function f assumed to have the Gaussian form

$$p(f \mid \boldsymbol{X}) = N(0, \boldsymbol{C}), \tag{1}$$

where the covariance matrix $\boldsymbol{C}$ is obtained from a *covariance function* $C(\cdot, \cdot)$ as in

$$C_{ij} = C(x_i, x_j) = \mathbb{E}\{f(x_i)f(x_j)\}. \tag{2}$$

One proceeds by assuming a parametric form for $C(x, x')$, with the parameters of the covariance function corresponding to hyperparameters of the final Gaussian process regression model. A common choice is the *Matérn covariance function*

$$C(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\,|x - x'|}{\lambda} \right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}\,|x - x'|}{\lambda} \right), \tag{3}$$

where K_ν is a modified Bessel function of the second order, and (σ^2, λ, ν) is the vector of hyperparameters. Of these, σ^2

[1]Drive tests necessarily have limited coverage due to the economic costs involved, and being conducted using vehicles imposes geographic limits to the region that can possibly be covered by the drive.

We introduce our example data set early on, as we believe that the best approach is to explain problems with concrete case studies with real data, and thereby ground the subsequent discussion on machine learning methods to realistic applications in mobile wireless networks.

controls the marginal variance of the process, λ the correlation distance, and ν the differentiability of the sample functions. All of these can be estimated by the method of maximum likelihood, after which closed form solution is available for the prediction y as conditional expectation at an arbitrary location as the function of the available data. Further, also the *estimation variance* is known explicitly without need for cross-validation as for many other machine learning methods. This is a very powerful property as it allows for the lightweight inference of the reliability of results and can even be used at runtime to guide the drive test being conducted.

2) Temporal Prediction Problems

A second category of performance prediction problems we consider is related to the *expected future performance* experienced by a *given mobile device*[2]. In such a case our data set $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ is typically temporally ordered, and our desired prediction y corresponds to the estimated value of x_{n+h} for some time horizon h. Clearly temporal and spatial prediction problems are inherently related, and methods for the latter can be applied to the former as well. However, given the unique features of the one-dimensional timeline, specialized machine learning methods are often used in the time domain. Here of particular importance are methods enabling incorporating typical temporal features such as seasonality combined with temporal correlations in subseasonal time scales into the model selection process. Without such structure in place (as is the case for simple autoregressive models for example) the needed model dimensionality easily explodes if complex temporal patterns are present in the data.

We consider two distinct approaches in the following. First is a state of the art Bayesian generalized additive model (GAM) style predictor [25], dubbed the *Prophet*, developed by the Facebook data science team and routinely used for business intelligence decisions in the applications of the company [26]. Second is the classical additive *Holt-Winters exponential smoothing algorithm* [27], [28] that applies exponential smoothing at different time scales to arrive at estimates of linear and periodic trends as well as for the final point predictions [29], [30]. The particular implementation of the Holt-Winters algorithm is documented in [31]. Of these, the former is undoubtedly the more powerful framework, with greater flexibility when it becomes to including different types of covariates (such as known changepoints in which the model is *expected* by the user to change its qualitative or quantitative behavior) into the prediction. However, flexibility always comes at a cost of *robustness*, especially when automated parameter estimation is involved. Our objective here is, therefore, to evaluate whether the more general framework results in significant enough improvements in prediction quality so as to warrant some loss in robustness.

3) Multidimensional Prediction Problems

Finally, the most general prediction problem we consider is the classical *multidimensional regression*. Here the objective is to predict the value of one selected variable in the data set using a general subset of all the others. This prevents us in general from exploiting the special structure spatial and temporal problems have, even though both spatial and temporal variables (without the understanding of their special role) can be incorporated at ease. By far the simplest, almost trivial example of a multidimensional regression algorithm is the *linear regression* method (LM). LM simply models y as a linear function of the covariates, as in

$$y = a_0 + \sum_i a_i x_i, \tag{4}$$

where coefficients a_i are determined based on the training data for example by minimizing the root mean squared error (RMSE) of the predictor. While obviously extremely simple, linear regression has in our context a simple communication-theoretic interpretation: in the high C/I regime when predicting the achieved user data rate, linear functions approximate well the Shannon capacity formula, and y becomes simply the best approximation of the network throughput as the optimal weighted sum of the individual Shannon capacity estimates. Thus, linear regression can be used as an improved proxy for simple SINR-based user data rate models relying on Shannon capacity formula.

A much more general and powerful family of regression techniques is obtained by considering *trees* of individual regression models. The model corresponds to a tree graph, with each nonleaf vertex corresponding to choosing a subspace by imposing an inequality of some of the x_i. The leaves of the tree finally yield the predictions y as the function of the ancestor vertices partitioning the space of x_i into subsequently finer subspaces. The various regression tree algorithms proposed in the literature differ mainly in the method used to choose the partitioning in terms of the covariates x_i, as well as in the way training data is used to find the optimum selection of decision variables in terms of the chosen partitioning scheme. We adopt here the *random forest* approach of Breiman

[2]This is a special case of independent interest of the more general spatio-temporal prediction problem for the entire radio environment map, allowing specialized techniques used for enhanced prediction results. The solutions for the general spatio-temporal case tend to essentially follow Gaussian process regression approaches, simply treating time as one additional dimension, thus adding no strong methodological insight [22].

[32] as used in [33]. Random forests have traditionally benefited from excellent prediction performance in diverse problems, and have, as we shall argue in Section V, other qualitative benefits that are not possessed by the vast majority of "general" machine learning algorithms.

Finally, we also explore the prediction performance of neural network models that have recently attracted considerable attention in the literature and in applications. Much of this attention has focused on the so-called "deep" architectures [34], although several alternative approaches exist as well. We consider here the *extreme learning machine* [13] for solving multidimensional regression problems as an example of a powerful class of neural networks which can be trained very rapidly, making them attractive for especially online applications, and therefore a commonly used initial step into more complex neural network applications[3]. Extreme learning machines are essentially feed-forward neural networks with a hidden layer of tunable size and optimizable activation function for each neuron. We allow the number of neurons in the hidden layer to be flexible, finding the optimal configuration for our training data through cross-validation.

Obviously, almost any machine learning method can be applied to multidimensional prediction problems. In our experience, it is critical to pay attention to the *understandability* of the model in addition to prediction performance. Trees and random forests have a structure that is easy for human operators in the loop to study and audit (we shall give an example of this in Section V), whereas for complex neural networks, for example, it is often difficult in networking applications to interpret the structure of the model in understandable terms.

[3]It is currently an open research question whether deep neural networks provide benefits for the applications considered in this article compared to simpler approaches considered here, mainly due to the sparsity of the drive test data making it difficult to correctly train complex machine learning architectures systems on it.

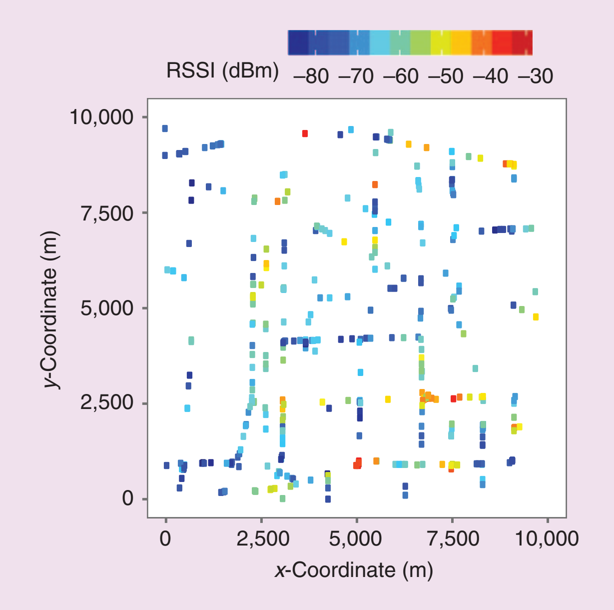

FIGURE 2 A random sample of 500 measurements used for spatial prediction at the rest of the measurement locations.

III. Spatial Prediction of Performance

We begin by studying the performance of Gaussian process regression on spatial prediction problems. Our example training data set consists initially of 500 measurements randomly selected from the entire test drive as illustrated in Figure 2 (we also discuss prediction of other variables from the data set, using the received signal strength values only as one example). We use the Matérn covariance model introduced above and find the hyperparameters from data by the method of maximum likelihood.

Figure 3 shows the arising prediction results for the entire $10\text{ km} \times 10\text{ km}$ example measurement area. In the figure we illustrate the final point predictions (Figure 3a), the theoretical

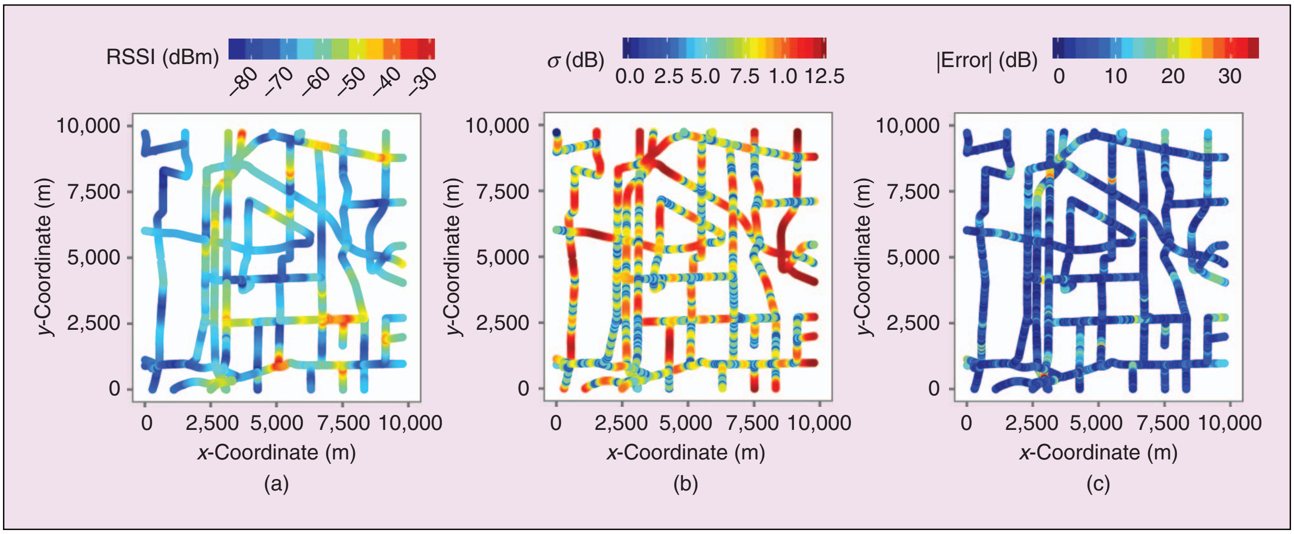

FIGURE 3 Spatial prediction of received signal strength using Gaussian process regression based on the sample of 500 measurements shown in Figure 2. (a) Prediction result, (b) Estimated σ, (c) Actual error.

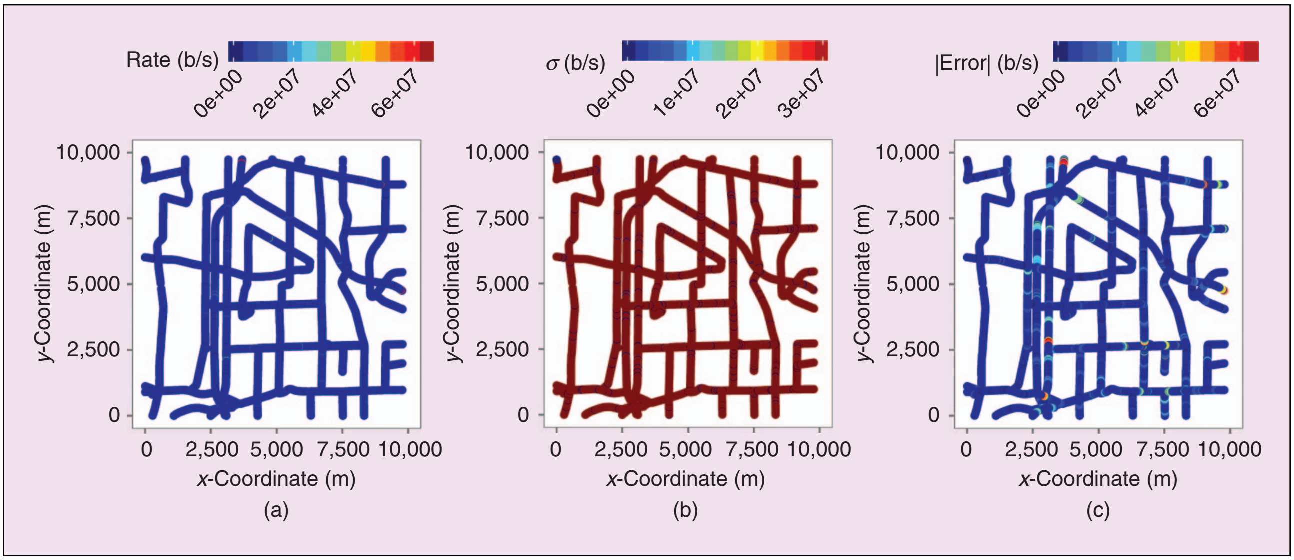

FIGURE 4 Spatial prediction of data rate using Gaussian process regression based on the sample of 500 measurements at locations shown in Figure 2. (a) Prediction result, (b) Estimated σ, (c) Actual error.

estimator variance expressed in terms of standard deviation σ (Figure 3b) and, finally, the actual absolute prediction error obtained by direct comparison against the ground truth (Figure 3c). Overall the prediction results are very good even though the measurement data set is fairly sparse (containing just 0.7% of the original data). The median absolute prediction error is only 2.5 dB, and even the 95th percentile is not excessive at 12.7 dB. The absolute prediction errors are somewhat heavy-tailed, with the mean of 3.95 dB significantly exceeding the median error. All of these performance figures indicate higher level of accuracy compared to modern predictive radio propagation models [35]–[37] even with adjustments for correlated shadow fading [38], [39]. We also see that the theoretical standard deviation results shown in Figure 3b are clearly pessimistic, indicating that they can be used as *conservative guidance* for performing drive tests as well as when evaluating the reliability of the estimation results.

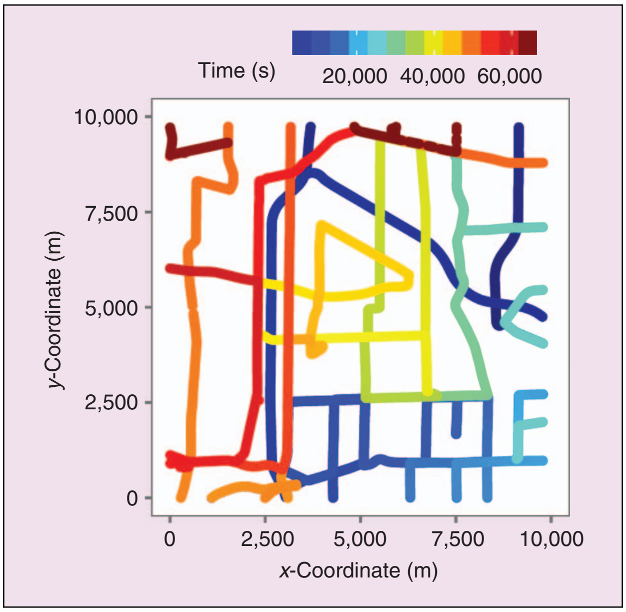

FIGURE 5 The time domain structure of the example drive test data set.

Figures 4a–4c show the analogous results in the case achieved user data rate is being used as the prediction target instead of received signal strength. As discussed in relation to Figure 1, the data rate varies much more erratically over the measurement region, making Gaussian process regression ineffective. Especially due to the small sample size the original training data misses several high data rate events and also does not lend itself to reliable estimation of the hyperparameters for the correlation function. Correspondingly the predictions closely resemble simple k-nearest neighbors predictions in structure, with the nearest measurement values essentially determining the prediction result. Fortunately for automated prediction applications, this prediction failure is readily apparent from the estimator variance depicted in Figure 4b, which is uniformly very high throughout the measurement region.

More generally, spatial estimation by Gaussian process regression tends to work well if the spatial sampling granularity is commensurate or finer compared spatial length scales of the quantity being predicted. In the case of received signal prediction, for example, the focus is typically on the local mean signal strength since this suffices for key applications such as coverage estimation or interference mapping for example, and multipath effects and other impairments with shorter length scales can be ignored.

IV. Temporal Prediction Problems

We shall now move on to illustrate temporal prediction problems related to our drive test data. Figure 5 shows the time domain structure of the drive through our example region (with measurements not falling into the measurement area removed). The mobile measurement platform gathers measurements at a rate of approximately one data point per second, indicating that the test drive through the region has taken close to 18 hours in total. Based on this data we consider the problem of predicting the next 30 seconds worth of measurements using the two machine learning methods introduced above, after an initial "burn-in"-period of 1000 seconds. We show all the results for a one-hour period following the initial burn-in time both in order to enable the reader to better discern the temporal structure involved, and also to avoid dealing with edge effects of the drive leaving briefly the measurement region. This one-hour period is contained in the dark blue road segments in Figure 5.

Figure 6 shows the prediction results for the received power, the C/I ratio, and the achieved user data rate using the "Prophet" predictor. Illustrated in the figures are the actual measurement readings (black dots), the mean 30-second forward rolling window prediction (red line) as well as the 95% confidence band around the latter (grey region around the mean prediction line). Overall the predictions are of adequate quality for the received power and the C/I ratio, with errors having standard deviations of 6.0 dB and 4.6 dB, respectively. However, from the figures, it is clear that Prophet somewhat oversmooths the data. This is especially apparent for the predictions of the data rate shown in Figure 6c, with the actual measurement data making frequent excursions well above the confidence bands.

Figure 7 shows the analogous prediction results for received power, the C/I ratio, and the achieved user data rate using the Holt-Winters exponential smoothing predictor instead. We see that, despite its apparent simplicity, for the prediction problems considered here, it seems to fare considerably better than the sophisticated Prophet does. This is confirmed by quantitative analysis of the prediction results, yielding standard deviations of 4.5 dB and 3.9 dB for the received power and the C/I ratio, respectively. Such performance is sufficient for a wide variety of optimization and management applications, in particular radio resource management based on prediction of future network conditions and demand instead of feedback-based reactive approaches commonly deployed in present-day networks. In particular, the range of C/I ratio values for typical cellular networking technologies varies over a range of 25 dB or more from cell-edge conditions to high quality of service near the base station. In that range, prediction accuracy of 3.9 dB gives already good indication of the radio conditions of the immediate future that can be exploited by the rest of the network for decision making.

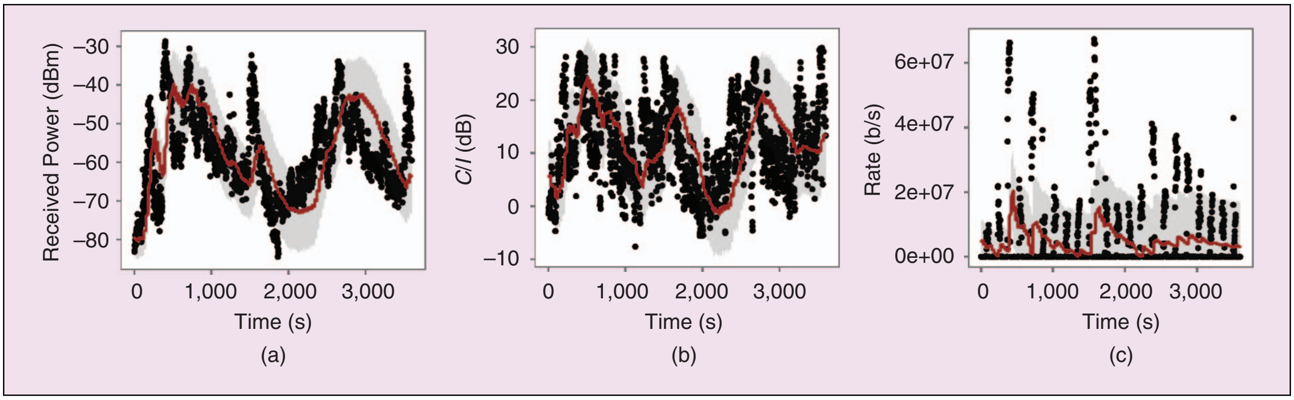

FIGURE 6 Time domain prediction results for a 30-second forward rolling window using the "Prophet" predictor. (a) Received signal strength, (b) *C/I* ratio, (c) Data rate.

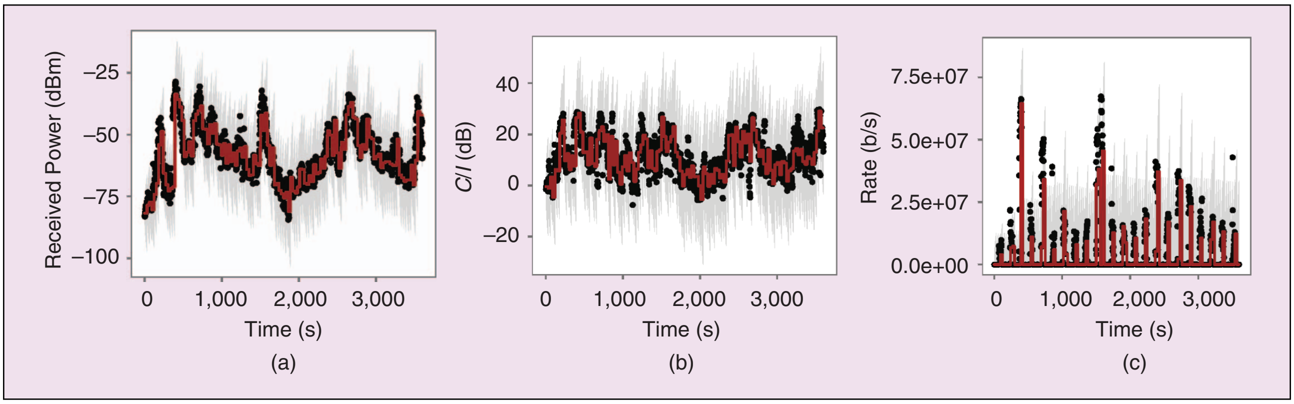

FIGURE 7 Time domain prediction results for a 30-second forward rolling window using Holt-Winters exponential smoother. (a) Received signal strength, (b) *C/I* ratio, (c) Data rate.

The range of C/I ratio values for typical cellular networking technologies varies over a range of 25 dB or more from cell-edge conditions to high quality of service near the base station. In that range prediction accuracy of 3.9 dB gives already good indication of the radio conditions of the immediate future that can be exploited by the rest of the network for decision making.

V. Multidimensional Prediction Problems

The third and most challenging problem type we introduced in Section II-B3 corresponds to the multidimensional prediction problem of trying to predict the value of one variable using others as predictors. The specific case we shall consider here is trying to predict the achievable user data rate based on either i) only RSSI, C/I ratio, and the terminal velocity, or ii) the previous covariates together with the absolute location of the user terminal. The latter types of models are not expected to generalize beyond the specific regions they were trained for, but can potentially result in higher prediction accuracy and yield insight into the root causes of performance problems in mobile networks. Such problems are especially interesting as traditionally estimates of achievable data rate can be determined only by actually setting up and using a data connection, whereas received signal strength and C/I ratio can be measured or estimated without transmitting data, and in much more rapid manner. This would be a significant step beyond the current state of the art.

Figure 8 shows the empirical CDF of the prediction error for linear regression, extreme learning machine neural network, and random forest based predictors for the case i) with only RSSI, C/I ratio, and the terminal velocity included as covariates. All methods were evaluated using half of the data set for hyperparameter estimation through repeated cross-validation and training, and second half of the data set for computing the realized prediction error for the trained model. As discussed in Section II-B3, we include linear regression as a baseline, yielding approximately the performance one would obtain from the commonly used theoretical models based on Shannon's capacity bound and similar results. As can be seen from the figure, the performance of these simplified models is rather poor due to the complexities of real-world mobile network operation. The neural network results are somewhat better, but the errors are still substantial even in the central area of the distribution. Random forest performs by far the best, with median prediction error of just 1 Mbps. However, even for the random forest case, the tail of the prediction error is very long, with extreme errors taking place at nonnegligible probability.

Figure 9 demonstrates results for the extreme learning machine for the case ii) of the problem, that is, when absolute spatial locations are included as covariates. Interestingly, there is no discernible impact when this additional information is included. This strongly indicates that the estimation error for the extreme learning machine does not have sufficiently simple structure spatially (for example, predictions failing at a single cell) that could easily be compensated by inclusion of additional location data as covariates.

Another major advantage of random forests and similar tree-based algorithms (main alternatives being boosted and bagged regression trees [20]) is that the final regression model is

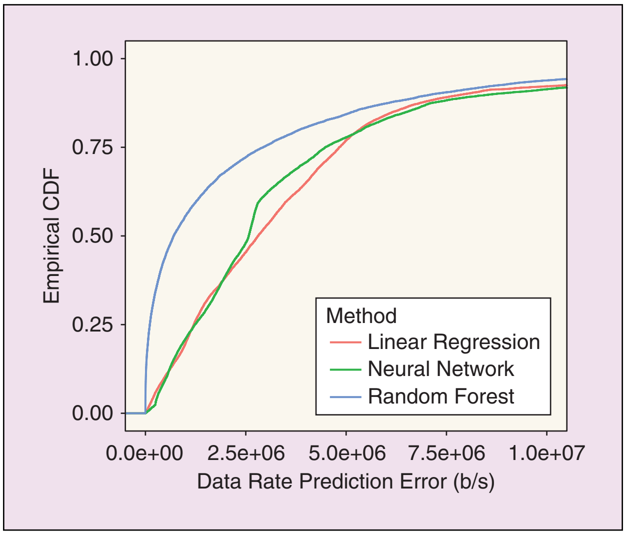

FIGURE 8 Performance of machine learning algorithms on multidimensional prediction of rate as function of received power, C/I ratio, and terminal velocity.

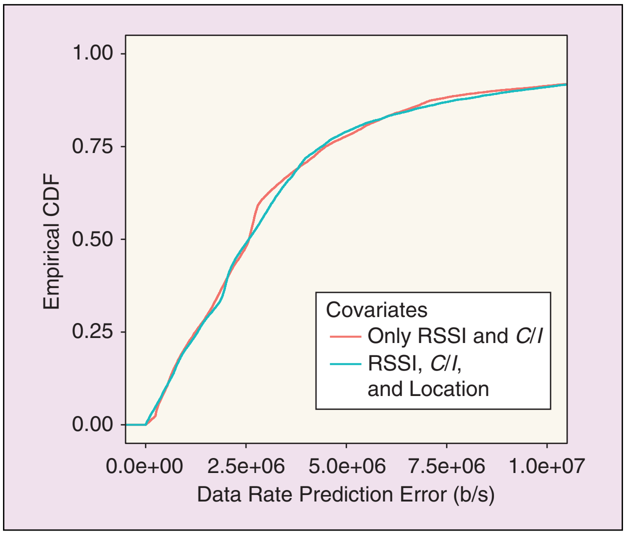

FIGURE 9 Illustration on the effect of including spatial data as covariates; due to model mismatch performance is not improved by including location data.

relatively understandable and in a sense transparent especially compared to deep learning architectures. Figure 10 illustrates one of the regression trees from the random forest model trained to obtain the results in Figure 8. The structure of the tree makes immediate intuitive and physical sense to the network engineer, with low C/I and RSS values resulting in low predicted data rates, and mobile nodes in general achieving lower data rates than almost stationary ones. An interesting research direction we are currently pursuing is to encode such intuition and experience in a formal *ontology* [12], which could be applied automatically to study the reasonability of a large number of regression trees arising in random forest models. Such checks can introduce an additional layer of robustness and trustworthiness into machine learning based automated decision making in wireless networks.

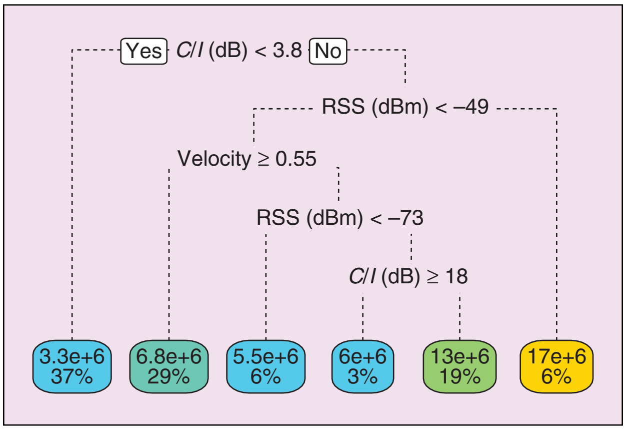

FIGURE 10 Example regression tree from the random forest for predicting data rate based on drive test data. Numbers in the leaf nodes at the bottom indicate the data rate prediction for this tree. The result for the entire random forest is based on the average of a large ensemble (in our case 500) of such regression tree results.

VI. Challenges for Future Work

Before concluding our paper we discuss challenges for future work that need to be overcome in order to make machine learning methods for performance prediction a practical component of operations and management of mobile wireless networks.

A. Architectural Issues

In the above examples we have focused on performance aspects of different machine learning techniques, applying them in an offline manner to a data set exported from a drive test system. In order to use such machine learning models for automatic decision making, optimization, and management of wireless networks they need to be integrated into the operations and management framework used by the network operator. This involves several system architecture design issues, including the harmonization of interfaces through which performance data is obtained [40]. Currently, several incompatible drive test systems are in use, creating a need for a unified interface and data formats abstracting away the underlying differences before machine learning models are applied. Such abstractions would also facilitate portability of already trained machine learning models between different systems.

B. Dealing with Severe Impairments in Data

Typical drive test data is of very high quality, collected using an expensive and carefully calibrated mobile measurement platform with static antenna placement usually at the rooftop of the measurement vehicle. As such drive tests are also very expensive to conduct, there has recently been substantial interest in complementing drive test data with measurements obtained through *crowdsourcing* [41], either through the standard *minimization of drive tests* interface [42], or through proprietary applications running on mobile terminals. Such measurements will suffer from a number of impairments, such as highly varying antenna placements, propagation effects due to the terminal being in the proximity of the human body, and various effects caused by nonmeasurement-grade hardware components used in typical user terminals. This necessitates development of preprocessing methodology to remove data with sufficiently severe impairments, and to develop machine learning models that can extract maximum of information from such measurements. We are currently working with *hierarchical Bayesian models* in an effort to model these impairments probabilistically and enable automated reasoning about them.

C. Scalability

The amount of data that can be gathered from mobile networks for performance prediction is vast, whereas many machine learning techniques become computationally infeasible for even moderately large data sets. Several tools are emerging from the "big data" research community to improve scalability of machine learning techniques, ranging from improved learning and inference algorithms to development of implementation infrastructures that enable massive parallelization of model training and applications [43]. Again random forests and similar models are well suited for such implementations, as the training process can be very efficiently parallelized.

D. Model Invalidation and Continuous Use

Classically machine learning is applied in a batch manner. A data set of fixed size is used to train the model, which is then applied for a period of time. Once the performance of the model begins to degrade, it is retrained with a newer data set. How to discover efficiently *changepoints* at which the system changes its behavior sufficiently to warrant retraining is a very important research topic for mobile networking applications. Another interesting research direction is the development of *online* models that can be continuously updated as new data becomes available [44].

E. Discretization for Robustness

All of our examples in this paper have been essentially regression estimation, where a numerical quantity is being predicted based on numerical covariates. Another major branch of machine learning—classification problems—also has great potential to be applied to mobile network optimization. The simplest way to apply classification systems for performance prediction is to discretize the numerical values used into categories that are meaningful to the application at hand. For example, user data rates could be discretized based on the bandwidth requirements of commonly used applications (e.g., video codecs). Often predicting performance in terms of such larger categories is more robust compared to regression for point estimates [45]. However, we are not aware of much especially empirical research of the effectiveness of such methods in applications discussed here.

VII. Conclusions

In this paper, we discussed the application of machine learning techniques for *performance prediction* problems in wireless networks. We studied the performance of existing machine learning algorithms for these problems and proposed a simple categorization of main problem types between spatial, temporal and multidimensional prediction tasks. Using an extensive real-world drive test data set, we showed that classical machine learning methods such as Gaussian process regression, exponential smoothing of time series, and random forests can yield very good prediction results for drive test data. We also discussed the key challenges for future work, especially with the focus of practical deployment of machine learning techniques for performance prediction in mobile wireless networks. We are currently working toward integrating prototype implementations of the proposed mechanism as a part of our radio environment map work [46].

References

[1] B. Bojović, E. Meshkova, N. Baldo, J. Riihijärvi, and M. Petrova, "Machine learning-based dynamic frequency and bandwidth allocation in self-organized LTE dense small cell deployments," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 183, pp. 16, Dec. 2016.

[2] R. Razavi, S. Klein, and H. Claussen, "Self-optimization of capacity and coverage in LTE networks using a fuzzy reinforcement learning approach," in *Proc. IEEE Int. Symp. Personal Indoor Mobile Radio Communications*, Istanbul, Turkey, Sept. 2010, pp. 1865–1870.

[3] S. Deb and P. Monogioudis, "Learning-based uplink interference management in 4G LTE cellular systems," *IEEE/ACM Trans. Netw.*, vol. 23, no. 2, pp. 398–411, Apr. 2015.

[4] A. Attar, V. Krishnamurthy, and O. N. Gharehshiran, "Interference management using cognitive base-stations for UMTS LTE," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 152–159, Aug. 2011.

[5] O. Sallent, J. Pérez-Romero, R. Ferrús, and R. Agustí, "Learning-based coexistence for LTE operation in unlicensed bands," in *Proc. IEEE Int. Conf. Communications Workshops*, London, U.K., June 2015, pp. 2307–2313.

[6] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Comput. Secur.*, vol. 28, no. 1–2, pp. 18–28, Feb.–Mar. 2009.

[7] S. Shamshirband, N. B. Anuar, M. L. M. Kiah, and A. Patel, "An appraisal and design of a multi-agent system based cooperative wireless intrusion detection computational intelligence technique," *Eng. Appl. Artif. Intell.*, vol. 26, no. 9, pp. 2105–2127, Oct. 2013.

[8] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Commun. Surv. Tutorial*, vol. 16, no. 4, pp. 1996–2018, Apr. 2014.

[9] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Anomaly detection in wireless sensor networks," *IEEE Wireless Commun.*, vol. 15, no. 4, pp. 34–40, Aug. 2008.

[10] D. Janakiram, V. Reddy, and A. P. Kumar, "Outlier detection in wireless sensor networks using Bayesian belief networks," in *Proc. Communication System Software Middleware*, New Delhi, India, Jan. 2006, pp. 1–6.

[11] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, KTH Royal Inst. Technol., 2000.

[12] S.-H. Liao, "Expert system methodologies and applications: A decade review from 1995 to 2004," *Expert Syst. Appl.*, vol. 28, no. 1, pp. 93–103, Jan. 2005.

[13] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.

[14] C. Clancy, J. Hecker, E. Stuntebeck, and T. O'Shea, "Applications of machine learning to cognitive radio networks," *IEEE Wireless Commun.*, vol. 14, no. 4, pp. 47–52, Aug. 2007.

[15] P. Mähönen, M. Petrova, J. Riihijärvi, and M. Wellens, "Cognitive wireless networks: Your network just became a teenager," in *Proc. IEEE INFOCOM*, Barcelona, Spain, Apr. 2006.

[16] R. W. Thomas, D. H. Friend, L. A. Dasilva, and A. B. Mackenzie, "Cognitive networks: Adaptation and learning to achieve end-to-end performance objectives," *IEEE Commun. Mag.*, vol. 44, no. 12, pp. 51–57, Dec. 2006.

[17] M. Zorzi, A. Zanella, A. Testolin, M. D. F. De Grazia, and M. Zorzi, "Cognition-based networks: A new perspective on network optimization using learning and distributed intelligence," *IEEE Access*, vol. 3, pp. 1512–1530, Aug. 2015.

[18] J. van de Beek, T. Cai, S. Grimoud, B. Sayrac, P. Mähönen, J. Nasreddine, and J. Riihijärvi, "How a layered REM architecture brings cognition to today's mobile networks," *IEEE Wireless Commun.*, vol. 19, no. 4, pp. 17–24, Aug. 2012.

[19] J. Perez-Romero, A. Zalonis, L. Boukhatem, A. Kliks, K. Koutlia, N. Dimitriou, and R. Kurda, "On the use of radio environment maps for interference management in heterogeneous networks," *IEEE Commun. Mag.*, vol. 53, no. 8, pp. 184–191, Aug. 2015.

[20] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*. Springer Series in Statistics, vol. 1. Springer: Berlin, 2001.

[21] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[22] N. Cressie and C. K. Wikle, *Statistics for Spatio-Temporal Data*. New York: John Wiley, 2015.

[23] A. Achtzehn, J. Riihijärvi, and P. Mähönen, "Improving accuracy for TVWS geolocation databases: Results from measurement-driven estimation approaches," in *Proc. IEEE DySPAN*, McLean, VA, Apr. 2014, pp. 392–403.

[24] H. Braham, S. B. Jemaa, G. Fort, E. Moulines, and B. Sayrac, "Fixed rank kriging for cellular coverage analysis," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4212–4222, May 2017.

[25] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*. Boca Raton, FL: CRC Press, 1990.

[26] S. J. Taylor and B. Letham. (2017). Forecasting at scale. Facebook Research Tech. Rep. [Online]. Available: https://facebookincubator.github.io/prophet/static/prophet_paper_ 20170113.pdf

[27] C. C. Holt, "Forecasting seasonals and trends by exponentially weighted moving averages," DTIC, Tech. Rep. AD0127522, Apr. 1957.

[28] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Manage. Sci.*, vol. 6, no. 3, pp. 324–342, Apr. 1960.

[29] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting Methods and Applications*. New York: Wiley, 2008.

[30] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," *Int. J. Forecasting*, vol. 18, no. 3, pp. 439–454, July–Sept. 2002.

[31] R. J. Hyndman and Y. Khandakar, "Automatic time series for forecasting: The forecast package for R," *J. Statist. Softw.*, vol. 27, no. 3, pp. 1–22, July 2008.

[32] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[33] A. Liaw and M. Wiener, "Classification and regression by RandomForest," *R. News*, vol. 2, no. 3, pp. 18–22, Dec. 2002.

[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[35] C. Phillips, D. Sicker, and D. Grunwald, "Bounding the error of path loss models," in *Proc. IEEE DySPAN*, Aachen, Germany, May 2011, pp. 71–82.

[36] C. Phillips, D. Sicker, and D. Grunwald, "A survey of wireless path loss prediction and coverage mapping methods," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 1, pp. 255–270, Mar. 2013.

[37] A. Molisch, L. Greenstein, and M. Shafi, "Propagation issues for cognitive radio," *Proc. IEEE*, vol. 97, no. 5, pp. 787–804, May 2009.

[38] J. Nasreddine, J. Riihijärvi, and P. Mähönen, "Transmit power control for secondary use in environments with correlated shadowing," in *Proc. IEEE Int. Conf. Communications*, Kyoto, Japan, June 2011.

[39] R. Fraile, J. Monserrat, N. Cardona, and J. Nasreddine, "Impact of slow fading modelling on TD-CDMA system-level simulations," in *Proc. Int. Symp. Wireless Communication Systems*, Valencia, Spain, Sept. 2006.

[40] M. Sooriyabandara, T. Farnham, P. Mähönen, M. Petrova, J. Riihijärvi, and Z. Wang, "Generic interface architecture supporting cognitive resource management in future wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 9, pp. 103–113, Sept. 2011.

[41] S. Rosen, S.-J. Lee, J. Lee, P. Congdon, Z. M. Mao, and K. Burden, "MCNet: Crowdsourcing wireless performance measurements through the eyes of mobile devices," *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 86–91, Oct. 2014.

[42] W. A. Hapsari, A. Umesh, M. Iwamura, M. Tomala, B. Gyula, and B. Sebire, "Minimization of drive tests solution in 3GPP," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 28–36, June 2012.

[43] T. Condie, P. Mineiro, N. Polyzotis, and M. Weimer, "Machine learning on big data," in *Proc. IEEE Int. Conf. Data Engineering*, Brisbane, Australia, Apr. 2013, pp. 1242–1244.

[44] S. Shalev-Shwartz, "Online learning and online convex optimization," *Found. Trends Mach. Learn.*, vol. 4, no. 2, pp. 107–194, Feb. 2012.

[45] S. Fidler, D. Skocaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 337–350, Mar. 2006.

[46] A. Achtzehn, J. Riihijärvi, I. A. Barriía Castillo, M. Petrova, and P. Mähönen, "CrowdREM: Harnessing the power of the mobile crowd for flexible wireless network monitoring," in *Proc. ACM HotMobile Conf.*, Santa Fe, NM, Feb. 2015, pp. 63–68.