

Article

Support Vector Machine-Based Transmit Antenna Allocation for Multiuser Communication Systems

Huifa Lin ¹, Won-Yong Shin ²  and Jingon Joung ^{3,*} 

¹ Telecommunication and Image Technology Laboratories, Sharp Corporation, Chiba 261-8520, Japan; huifa.lin.dr@ieee.org

² Computational Science and Engineering, Yonsei University, Seoul 03722, Korea; wy.shin@yonsei.ac.kr

³ School of Electrical and Electronics Engineering, Chung-Ang University, Seoul 06974, Korea

* Correspondence: jgjoung@cau.ac.kr; Tel.: +82-2-820-5145

Received: 4 April 2019; Accepted: 3 May 2019; Published: 6 May 2019



Abstract: In this paper, a support vector machine (SVM) technique has been applied to an antenna allocation system with multiple antennas in multiuser downlink communications. Here, only the channel magnitude information is available at the transmitter. Thus, a subset of transmit antennas that can reduce multiuser interference is selected based on such partial channel state information to support multiple users. For training, we generate the feature vectors by fully utilizing the characteristics of the interference-limited setup in the multiuser downlink system and determine the corresponding class label by evaluating a key performance indicator, i.e., sum rate in multiuser communications. Using test channels, we evaluate the performance of our antenna allocation system invoking the SVM-based allocation and optimization-based allocation, in terms of sum-rate performance and computational complexity. Rigorous testing allowed for a comparison of a SVM algorithm design between one-vs-one (OVO) and one-vs-all (OVA) strategies and a kernel function: (i) OVA is preferable to OVO since OVA can achieve almost the same sum rate as OVO with significantly reduced computational complexity, (ii) a Gaussian function is a good choice as the kernel function for the SVM, and (iii) the variance (kernel scale) and penalty parameter (box constraint) of an SVM kernel function are determined by 21.56 and 7.67, respectively. Further simulation results revealed that the designed SVM-based approach can remarkably reduce the time complexity compared to a traditional optimization-based approach, at the cost of marginal sum rate degradation. Our proposed framework offers some important insights for intelligently combining machine learning techniques and multiuser wireless communications.

Keywords: antenna allocation systems; multiuser communication systems; multiclass classification; supervised machine learning; support vector machine

1. Introduction

Recently, machine learning has been attracting much research interest from various fields due to numerous successful applications to solve significant practical problems [1–11]. Most of the conventional approaches in communication system design rely on maximizing or minimizing the objective functions, i.e., optimization-driven approaches. However, for some problems, one has to resort to algorithms with fast-increasing complexity, e.g., the antenna selection/allocation problem in multiuser communication systems [12–15]. Hence, for future application scenarios with large-scale configurations, such as massive multiple-input and multiple-output (MIMO) systems and machine learning-based methods, a data-driven approach seems to be more promising because it is possible to provide near-optimal communication performance with relatively low online prediction complexity, leaving the high complexity part to the offline training phase in machine learning. In the

research fields of communication and signal processing, different types of machine learning methods including (i) unsupervised learning, (ii) reinforcement learning, and (iii) supervised learning find suitable application areas in many areas [16–19]. In this study, we attempt to explore the possibility of applying supervised learning techniques (switching the optimization-driven thinking to the data-driven thinking) to solve a specific antenna allocation problem in multiuser communication systems. Even compared to the optimization-driven antenna selection algorithms with reduced complexity [13,20–25], our proposed supervised learning-based approach is a promising alternative since it demonstrates almost optimal performance while greatly reducing complexity. Moreover, our supervised learning-based approach can also provide design flexibility based on a trade-off between computational complexity and sum-rate performance by adjusting the number of effective classes, i.e., computational complexity can be further reduced by using a smaller number of effective classes, with tolerable sum-rate performance degradation.

1.1. Related Work

An early study investigated dynamic channel assignment for cellular systems by using neural networks [26]. One of the most popular unsupervised learning methods, k -means clustering, was used to jointly optimize both the gateway partitioning and the virtual-channel allocation in hybrid optical/wireless networks [27], to group distributed users in networks [28], and to blindly detect symbols [29]. Decision trees and neural networks were applied to enhance the routing performance [30].

Reinforcement learning techniques have also been found to be helpful in solving multi-armed bandit problems in wireless communications [31,32], link adaptation [33], and to autonomously manage heterogeneous networks [34]. Other reinforcement learning methods, such as Markov decision processes and Q-learning, have been applied to energy harvesting sensor networks [35], cognitive radio networks [36], and opportunistic access of macro/femtocell heterogeneous networks [37]. Accelerated reinforcement learning algorithms have been used to predict spectrum opportunities in opportunistic spectrum access networks, resulting in improved secondary sum-rate performance [38].

Supervised learning methods, such as the k -nearest neighbors algorithm (k -NN) and the support vector machine (SVM), have been employed to design link adaptation methods for wireless communication networks [39–41]. There have been extensive studies on not only improving the accuracy or complexity of SVMs [42], but also applying SVM to various classification fields such as signal detection in visible light communications [43]. The efficiency of k -NN-based beam allocation [44] was also demonstrated in multiuser massive MIMO systems. The whole communication system was interpreted as a machine learning system, i.e., mapping major components in the communication system including the transmitters, channels, and receivers to different layers in a neural network [45,46]. In [47], the communication system was considered as an autoencoder and the optimization of the communication system was accomplished by training the corresponding neural network.

Among the attempts to apply machine learning to the communication and signal processing fields, and specifically to the antenna allocation problem considered here, a recent work has applied supervised machine learning-based methods for antenna allocation in multi-antenna point-to-point (single-user) wireless networks [48]. It was shown that the antenna allocation in wireless communications can be converted to a multiclass classification problem. Compared to the conventional optimization-driven methods in [12–14], the multiclass classification, learning-based method in [48] provided near-optimal communication performance with relatively low online computational complexity.

It is worth noting that computational complexity is one of the bottlenecks for traditional optimization-driven antenna selection approaches because of the rapidly increasing complexity of exhaustive search algorithms for the larger number of antennas [13,20]. To reduce computational complexity, seminal studies have proposed optimization-driven antenna selection methods in single-user communication systems [21,22]. For multiuser communication systems, there have also been many studies attempting to reduce the complexity of antenna selection [23–25].

1.2. Motivations and Contributions

Inspired by the pioneering work on learning-based antenna allocation for a single-user system [48], we focus on a wireless downlink multiuser communication system with antenna allocation. Extending the single-user system to multiuser networks, it is important to properly determine one or more communication performance metrics, such as sum rate and bit error rate (BER) [49]. For multiuser communication systems, maximizing the system sum rate might be a proper design target rather than minimizing the BER, due to the presence of interference among multiple users. Moreover, it is easy to obtain a rather trivial solution to the BER minimization problem for a multiuser communication system by selecting only one user (but not multiple users). In [41], learning-based link adaptation methods were designed for multiuser communication networks, where a greedy algorithm was employed for user selection. However, rather than improving fairness by user scheduling, our aim is to maximize the sum rate of all the users in the network by allocating the proper antenna to each user. Hence, we first formulate antenna allocation in the multiuser network as a multiuser sum rate maximization problem and solve it by using multiclass classification algorithms. To the best of our knowledge, this study is the first attempt at machine learning-based antenna allocation for multiuser wireless networks.

The main contributions of this study are summarized as follows:

- We interpret the antenna allocation system for multiuser communication as a multiclass classification learning system. For the components of the learning system, such as the training data and the corresponding class labels, we first model the counterparts in the conventional communication system, and we then construct them with a proper format for the learning system.
- We establish a communication system with an SVM module that allocates transmit antennas to each user in multiuser communication networks with partial channel state information at the transmitter (CSIT), as shown in Figure 1. The antenna allocation method is designed for a frequency-division duplex (FDD) system, where only quantized channel gain information is available.
- The parameters of the designed SVM are tuned based on extensive numerical experiments in order to improve the sum-rate performance of the communication system. For the designed SVM, we find that a Gaussian function is a good choice for the kernel function, which is one of the most important parameters for tuning SVMs. (Artificial neural networks (ANNs) can also be employed in our learning system, which may result in a slightly better sum-rate performance at the expense of both higher computational complexity and a larger training dataset than for k -NN and SVM. Thus, ANNs are not considered in this study as our main focus is on the design of learning systems showing a significant reduction in complexity from the optimization-based approach with marginal sum rate reduction.) From our rigorous simulation, the variance (kernel scale) and penalty parameter (box constraint) of an SVM kernel function are determined by 21.56 and 7.67, respectively.
- Numerical experiments are extensively performed for various configurations of communication systems in order to evaluate the proposed SVM-based antenna allocation method. We find that with lower online computational complexity, the designed SVM method achieves near-optimal performance, as is obtained from the conventional optimization approach. Compared to the k -NN method, the SVM method is superior not only in terms of sum-rate performance but also prediction complexity performance.

1.3. Organization

The rest of the paper is organized as follows. In Section 2, the system model, optimization problem formulation, and corresponding optimization-driven solutions are described. In Section 3, we introduce the proposed machine learning-based antenna allocation method from overall framework structure to implementation details. In Section 4, the proposed method is evaluated to determine the parameters of the designed learning system, and its performance is compared to that of the conventional optimization-driven method. Finally, we conclude this paper in Section 5.

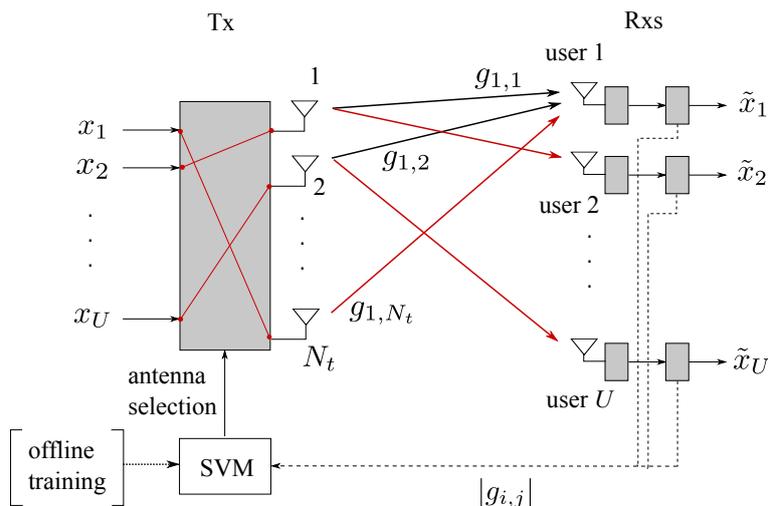


Figure 1. System model of the considered multiuser communication network consisting of one transmitter with N_t antennas and U users with a single antenna.

2. System Model and Conventional Optimization Approach

2.1. System Model

As illustrated in Figure 1, we consider a multiuser communication network consisting of one multi-antenna transmitter (Tx) and U selected/scheduled single-antenna receivers (Rxs or users). Let $N_t \geq U$ denote the number of antennas at the Tx. We use $\mathbf{h}_i \in \mathbb{C}^{N_t \times 1}$ to denote the channel coefficient vector from the transmitter to receiver i , where $i \in \{1, 2, \dots, U\}$. Let $\mathbf{H} \in \mathbb{C}^{U \times N_t}$ be an overall channel coefficient matrix from the transmitter to all the receivers, where the i th row of \mathbf{H} is given by \mathbf{h}_i^T and the (i, j) th entry of \mathbf{H} is expressed as $h_{i,j}$, that is, the channel from Tx antenna j to user i . In this study, we consider a low-cost Tx that has a limited number of radio frequency (RF) chains and computational capability. Thus, an antenna selection scheme using a part of transmit antennas is relevant, rather than the highly complex optimal multiuser pre-processing or beamforming schemes. Accordingly, channel magnitude information, i.e., partial CSI, is sufficient for simple antenna selection at the Tx (refer to Section 3 for the overall procedure). Note that the main motivation of this study is to effectively reduce the computational complexity at the Tx by using a machine learning-based method. The partial CSIT is available through channel gain feedback from Rxs to Tx. Specifically, each Rx i estimates its own channel gains, i.e., $h_{i,j}$ for all j , obtains their magnitude information $g_{i,j} = |h_{i,j}|^2$, and feeds them back to the Tx. The Tx accumulates the feedback information from all Rxs and constructs a channel gain matrix $\mathbf{G} \in \mathbb{R}^{U \times N_t}$ whose real-value entries are given by $g_{i,j} = |h_{i,j}|^2$.

2.2. Optimization Problem Formulation

The Tx selects U antennas from N_t antennas to allocate one antenna to each user. Then, U independent data streams are transferred to the U users, i.e., data stream x_i is delivered to user i and is estimated by \tilde{x}_i . For a certain antenna allocation scheme indexed by l , let $s_i^{(l)}$ denote the index of the antenna allocated to user i , where $s_i^{(l)} \in \{1, 2, \dots, N_t\}$ and $s_i^{(l)} \neq s_j^{(l)}$ if $i \neq j$ ($i, j \in \{1, 2, \dots, U\}$). The corresponding index vector of the U allocated antennas is then defined as $\mathbf{s}_l \triangleq [s_1^{(l)}, s_2^{(l)}, \dots, s_U^{(l)}]^T$, where the superscript T represents the transpose of a vector or matrix, and index $l \in \mathcal{L}$ is used to denote the antenna allocation scheme. Here, we define $\mathcal{L} \triangleq \{1, 2, \dots, L\}$ as a set of antenna allocation schemes, where the number of all available \mathbf{s}_l (all valid antenna allocation schemes), L , is given by

$$L \triangleq \frac{N_t!}{(N_t - U)!}. \tag{1}$$

Now, suppose that the antenna allocation scheme is given by \mathbf{s}_l . We can then compute the resultant data rate of user i as follows [50]:

$$R_i(\mathbf{s}_l) = \log_2 \left(1 + \frac{P_{\text{TX}} g_{i,s_i}^{(l)}}{N_0 + \sum_{j \neq i}^U P_{\text{TX}} g_{i,s_j}^{(l)}} \right), \quad (2)$$

where P_{TX} is the transmit power, N_0 is the noise variance, $P_{\text{TX}} g_{i,s_i}^{(l)}$ is the power of the desired signal to user i , and $\sum_{j \neq i}^U P_{\text{TX}} g_{i,s_j}^{(l)}$ is the power of interference signals from the antennas allocated to other users. Then, the sum rate of the system for the antenna allocation \mathbf{s}_l is given by

$$R_{\text{sum}}(\mathbf{s}_l) = \sum_{i=1}^U R_i(\mathbf{s}_l). \quad (3)$$

The optimization problem is then formulated as follows:

$$\underset{l \in \mathcal{L}}{\text{maximize}} \quad R_{\text{sum}}(\mathbf{s}_l). \quad (4)$$

2.3. Optimization-Driven Solutions

To solve the formulated combinatorial optimization problem in Equation (4), a brute-force exhaustive search (or any other more sophisticated optimization-driven algorithms) with high computational complexity can be applied to find l^* that maximizes R_{sum} . First, the data rates of U users, i.e., $R_i(\mathbf{s}_l), i \in \{1, \dots, U\}$, are computed by using Equation (2) for all the L antenna allocation schemes. Then, the sum rate can be computed based on (3). Among the antenna allocation schemes, the optimal index l^* that maximizes the value of R_{sum} is determined. For each antenna allocation scheme, the exhaustive search algorithm traverses all U users to compute the data rate, where $(U - 1)$ computations are required to calculate the total interference term. Thus, the computational complexity of this exhaustive search is given by $\mathcal{O}(LU(U - 1)) = \mathcal{O}\left(\frac{U^2 N_l!}{(N_l - U)!}\right)$. It can be seen that the complexity increases and becomes prohibitively large as N_l or U increases.

For better readability, the main notations used here to describe communication systems are summarized in Table 1.

Table 1. Notations used to describe communication systems.

Notation	Description
N_l	number of antennas at the transmitter
U	number of users
\mathbf{h}_i	channel coefficient vector to device i
\mathbf{H}	overall channel coefficient matrix
\mathbf{G}	overall channel gain matrix
\mathbf{s}_l	index vector of the allocated antenna with label l
\mathcal{L}	set of labels for all the available antennas allocated
L	number of labels in \mathcal{L}
P_{TX}	transmit power per antenna
R_{sum}	sum rate of the system in bps/Hz

3. SVM-Based Antenna Allocation

In order to reduce the computational complexity of the exhaustive search for antenna selection, we consider SVM-based antenna selection to solve Equation (4). Specifically, we employ a multiclass SVM algorithm to classify channel gain samples into L classes, each of which corresponds to an available antenna allocation scheme. With a sufficient number of channel gain samples, i.e., training data, we can

design a classification model, which can be used to predict the class of a test channel gain matrix, i.e., the best antenna allocation scheme for a new channel realization in a test, i.e., actual communications.

Generally speaking, in a machine learning system, a learning model is first trained by the input training set and the corresponding labels [51,52]. Then, this learning model can be used to predict the class labels for a new test dataset. The overall machine learning framework of our antenna allocation system is illustrated in Figure 2. It is worth noting that there are two types of tasks in this framework, an offline task and an online task. The online task includes channel estimation and channel allocation through learning-based prediction. The offline task consists of three tasks: i) training sample set design, ii) learning systems design, and iii) parameter adjustment. Distinguishing the online and offline tasks is crucial in communication systems. This is because the offline task can be performed with more powerful computing resources and relaxed computational complexity requirements, while the online task typically has stringent latency and computing constraints. In the following three subsections, the three offline tasks will be described in detail.

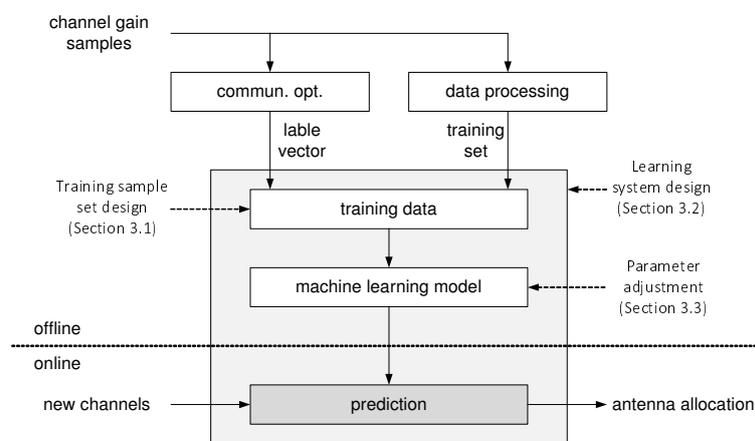


Figure 2. Machine learning framework for antenna allocation in a multiuser communication system.

3.1. Task 1: Designing a Training Sample Set

We need to manipulate the matrix form of the channel gain samples into training data with a suitable form for input into the learning system. Three procedures are performed to obtain the training data for the machine learning system (not necessarily in sequence): (i) design training data from the channel gain matrices, (ii) design the key performance indicator (KPI), and (iii) declare the corresponding label based on the KPI, i.e., labeling.

3.1.1. Subtask 1-1: KPI Design

A KPI is designed to label the training set. In general, a KPI can be defined as any metric used in communications, such as spectral efficiency, energy efficiency, BER, effective signal-to-noise ratio (SNR), communications latency, and any combination thereof [53]. In this study, we use the sum rate of a system, R_{sum} , as the target KPI.

3.1.2. Subtask 1-2. Training Set Design

The training samples are the input for a learning system and are known as input variables, predictors, or attributes. As shown in Section 2.1, we assume that magnitude channel information, i.e., the channel gain matrix \mathbf{G} , is available in our communication system. Based on the available channel information, it is important to properly design the training set by taking into account not only the target KPI, which affects communication performance, but also the complexity of the system. For example (refer to [54] and the references therein), singular values are used for a singular value-based antenna selection system, the minimum eigenvalues of the Hermitian matrix of the channel matrix for Gerschgorin circle-based antenna selection systems, channel norm values for norm-based antenna

selection systems, and the dot products of channel column vectors for correlation-based antenna selection systems. Here, the singular values are clearly a good candidate for a training set for various KPIs (e.g., spectral efficiency, energy efficiency, and BER), yet they require a higher complexity higher than the other values. In this study, we adopt a signal-to-interference-leakage ratio (SILR) metric that is closely coupled with the sum rate, which is our target KPI. It is worth noting that with this SILR metric, it is sufficient to perform antenna allocation with the channel gain knowledge at the Tx, even without knowing the signal-to-interference-plus-noise ratio (SINR) of the users. We also note that it may not be possible to acquire the received SINR at the Tx under our communication mechanism using the channel gain feedback. This is because the sum of interference links can be computed after the antenna allocation process and the set of allocated antennas is not available in the training phase.

An SILR matrix, $\mathbf{Z} \in \mathbb{R}^{U \times N_t}$, is employed to generate the training set, which can be computed based on \mathbf{G} . Specifically, the entry of the i th row and the j th column of \mathbf{Z} is given by

$$Z_{i,j} = \frac{g_{i,j}}{\sum_{\substack{k=1 \\ k \neq j}}^U g_{i,k}}. \quad (5)$$

From Equation (5), it can be seen that the SILR metric simultaneously captures both the desired signal strength and the interference leakage to other users. Because a machine learning system requires the real-value vector input of multiple features, we transform the SILR matrix into a $\mathbb{R}^{1 \times N_t U}$ vector by stacking U users' vectors (the row vectors in \mathbf{Z}). The training set vector with $N_t U$ features is given by

$$[Z_{1,1}, \dots, Z_{1,N_t}, \dots, \underbrace{Z_{i,1}, \dots, Z_{i,N_t}}_{\text{the } i\text{th row of } \mathbf{Z}}, \dots, Z_{U,1}, \dots, Z_{U,N_t}]. \quad (6)$$

By repeating the channel generation and data processing D times, we obtain a training set matrix $\mathbf{T}_{\text{raw}} \in \mathbb{R}^{D \times N_t U}$ whose rows are given by the training vector in Equation (6). As a special case, when $U = 1$ (single-user communication systems), we use the channel gain as the metric (i.e., $Z_{i,j} = g_{i,j}$) due to the absence of interferences. Since \mathbf{Z} is a vector for $U = 1$, we can skip the aforementioned stacking step by directly using \mathbf{Z} as the training set vector.

The final step of training set design is to normalize the training samples to obtain a proper input training set for the learning system. Let $\mathbf{T} \in \mathbb{R}^{D \times N_t U}$ denote a training set matrix as one of the inputs to the learning system, whose (i, j) th element, denoted by $T_{i,j}$, is a normalized value of the (i, j) th element of \mathbf{T}_{raw} .

$$T_{i,j} = \frac{Z_{i,j} - \mathbb{E}_i(Z_{i,j})}{\max_i(Z_{i,j}) - \min_i(Z_{i,j})}, \quad (7)$$

where the term $\max_i(Z_{i,j}) - \min_i(Z_{i,j})$ in (7) indicates the normalization used for improving the learning speed/convergence or avoiding a precision issue with very large- or small-value data.

3.1.3. Subtask 1-3. Class Design and Labeling

From the interpretation of the antenna allocation process and multiclass classification, it is clear that designing the labeling is equivalent to designing the antenna allocation scheme. As shown in Section 2.2, the mapping from antenna allocation \mathbf{s}_l to the index l is a one-to-one mapping. Thus, we can use the index set \mathcal{L} for the labeling in the machine learning system. Let $\mathbf{c} = [c_1, \dots, c_D]^T$ denote the class label vector for the training data matrix \mathbf{T} , where $c_i \in \mathcal{L}$ and $i \in \{1, \dots, D\}$. Thus, we use this metric in the following content. The labeling procedure is summarized as follows:

- Evaluate the target KPI, η , for the d th channel gain sample with a particular antenna allocation \mathbf{s}_l corresponding to label $l \in \mathcal{L}$.
- Assign the d th element of \mathbf{c} , c_d , with l^* , which stands for the best choice among all the antenna allocation schemes.
- Repeat the previous two steps for D times to go through all D training set.

Remark on the reduction of the number of classes: We can further improve the labeling by exploiting the knowledge of a wireless communication system. It is known that multiple antennas with less spatial correlation results in better communication performance, which is also confirmed by our numerical experiments. Therefore, we reduce the number of classes to $L' < L$ by deleting some less selected classes, which correspond to the schemes with highly correlated antennas. This elimination can reduce prediction complexity, with a tradeoff in classification performance. Note that even with classes that are uniformly selected, a designer can still reduce L to L' to reduce the complexity of the learning system if the resultant performance degradation is marginal. On the other hand, the number of clusters can also be automatically determined for unsupervised clustering systems by using Davies-Bouldin or Dunn indices (refer to [55] and references therein).

3.2. Task 2: Designing Learning Systems

From Task 1, we obtain the real-value matrix $\mathbf{T} \in \mathbb{R}^{D \times N_t U}$ as the training set and the corresponding class label vector $\mathbf{c} = [c_1, \dots, c_D]^T$. Using the labeled training dataset (training data), i.e., \mathbf{T} and \mathbf{c} , we build a learning system, and specifically, a trained multiclass classifier whose input is an estimated channel gain vector and whose output is the index of the antenna allocation scheme. Since $L > 2$ in our antenna allocation system, we employ L -class classification algorithms, such as the multiclass k -NN and SVM algorithms. For the simple description of the multiclass classification algorithms, we denote the i th row vector of \mathbf{T} by $\mathbf{t}_i \in \mathbb{R}^{1 \times N_t U}$.

We now introduce the fundamental mechanism of a binary SVM classifier and then explain how to perform multiclass classification based on the binary SVM classifier. With a binary SVM, the data are separated into two half-spaces with a hyperplane $f(\mathbf{t})$, which is given by

$$f(\mathbf{t}) = \mathbf{w}\mathbf{t}^T + \beta = 0, \quad (8)$$

where $\mathbf{t} \in \mathbb{R}^{1 \times N_t U}$ is a feature vector, $\mathbf{w} \in \mathbb{R}^{1 \times N_t U}$ is a weight vector, and $\beta \in \mathbb{R}$ is a biasing variable. Here, the linear kernel function, denoted by $K(\mathbf{t}_i, \mathbf{t}_j) = \mathbf{t}_i^T \mathbf{t}_j$, is employed, but generally, various types of kernel functions can also be adopted. This will be discussed later in this section. A classification rule induced by $f(\mathbf{t})$ for the new observation \mathbf{t}_{new} is

$$\text{sign}(\mathbf{w}\mathbf{t}_{\text{new}}^T + \beta). \quad (9)$$

The training data samples that are nearest to the decision boundary are called support vectors, where the distance is given by $\frac{1}{\|\mathbf{w}\|_2}$. Thus, in order to separate the data as much as possible, the margin that is given by $\frac{2}{\|\mathbf{w}\|_2}$ needs to be maximized. This optimization problem is equivalent to minimizing $\|\mathbf{w}\|_2$. Since the training data may be not totally separable (that is the case for our antenna allocation system), the optimization problem is formulated by introducing slack variable $\xi_i \geq 0, \forall i = \{1, 2, \dots, N_t U\}$ as

$$\begin{aligned} & \underset{\mathbf{w}, \beta}{\text{minimize}} && \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^{N_t U} \xi_i \\ & \text{subject to} && \xi_i \geq 0, c_i (\mathbf{w}\mathbf{t}_i^T + \beta) \geq 1 - \xi_i \quad \forall i, \end{aligned} \quad (10)$$

where the ‘‘penalty’’ parameter C is used to penalize the training error of the soft margin SVM. This parameter C needs to be tuned for good classification performance because too large a C causes overfitting and too small a C causes underfitting. The decision boundary can be found by solving this convex quadratic optimization problem. As mentioned before, by adopting different kernel functions, we can apply the ‘‘kernel trick’’ to map the original feature space to a higher-dimensional feature space where the training set could be more separable. For instance, the polynomial kernel function and the Gaussian kernel function are given by

$$K(\mathbf{t}_i, \mathbf{t}_j) = \left(1 + \mathbf{t}_i^T \mathbf{t}_j\right)^p \quad (11)$$

and

$$K(\mathbf{t}_i, \mathbf{t}_j) = e^{-\frac{\|\mathbf{t}_i - \mathbf{t}_j\|^2}{2\sigma^2}}, \quad (12)$$

respectively, where $p \geq 2$ is the polynomial power and σ^2 is the variance. We can do this because the optimization process of SVM allows us to simply modify the kernel function $K(\mathbf{t}_i, \mathbf{t}_j)$ by replacing the linear kernel function with other kernel functions, without changing the overall optimization algorithm.

In order to perform multiclassification using the binary SVM method that was originally designed for binary classification [56], we can either employ the one-vs-all (OVA) strategy with L binary SVM learners or the one-vs-one (OVO) strategy with $\frac{L(L-1)}{2}$ binary SVM learners [57]. Compared to the OVO strategy, the OVA strategy has a lower computational complexity because fewer binary SVM learners are required. The selection of the proper multiclassification strategy and kernel function will be discussed in detail based on numerical results in the next subsection.

3.3. Task 3. Parameter Adjustment

For the multiclass SVM algorithms, several parameters and strategies can be selected and tuned in order to achieve superior sum-rate performance. Here, we note that machine learning algorithms with low prediction complexity are favorable to antenna selection in communication systems. Based on numerical results regarding the sum-rate performance and prediction runtime performance, we discuss the selection of kernel function from linear, polynomial, and Gaussian kernel functions, and the selection of multiclass strategy from between OVA and OVO. Through rigorous simulation and comparison, we find that the SVM algorithm with the OVA strategy and the Gaussian kernel provides nearly the best sum-rate performance with a relatively low prediction runtime, and is thus adopted in our antenna selection system. For example, Figures 3 and 4 verify our observation when $N_t = 5$ and $U = 3$, where solid red curves represent the use of SVM algorithms with the OVA strategy while dashed blue curves represent the use of SVM algorithms with the OVO strategy. In contrast, SVM algorithms with the OVO algorithm show a relatively high prediction runtime that grows fast with L' , although the sum-rate performance is slightly superior to that of the OVA counterparts. The SVM algorithm with the OVA strategy and linear kernel provides the lowest prediction runtime performance but the sum-rate performance degradation is severe.

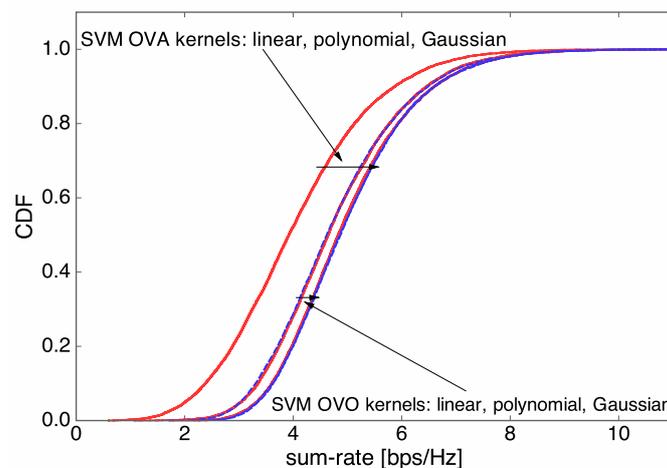


Figure 3. Support vector machine (SVM) performance evaluation results when $N_t = 5$ and $U = 3$: the empirical cumulative density function (CDF) of the sum rate for various SVM kernel functions.

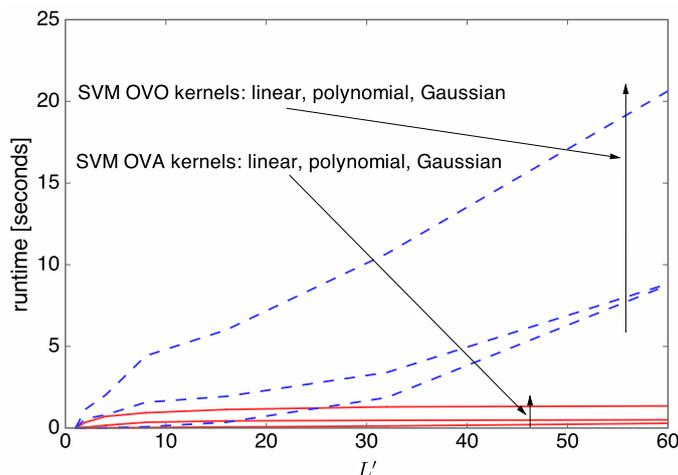


Figure 4. SVM performance evaluation results when $N_t = 5$ and $U = 3$: runtime over L' for various SVM kernel functions.

For each combination of multiclass strategies and kernel functions, proper values of crucial parameters such as the variance σ^2 and the penalty parameter C are found by extensive experimentation via a heuristic search targeting the minimum cross-validation loss. For instance, when the OVA strategy and the Gaussian kernel function are adopted, σ^2 and C are determined by 21.56 and 7.67, respectively.

4. Numerical Evaluation

In this section, we evaluate the performance of the designed communication system invoking SVM-based allocation with the OVA strategy and the Gaussian kernel function in terms of sum rate and computational complexity via computer simulations. For comparison, we also consider three benchmark systems: (i) OPT, which is an optimization-driven method that maximizes the sum rate, i.e., exhaustive search, discussed in Section 2.3, (ii) RAND, which selects antennas randomly, (iii) and k -NN, an antenna allocation system based on the k -NN algorithm instead of SVM. Unless otherwise stated, we take into account full permutations of an antenna set (i.e., full classes), e.g., $L = 60$ when $N_t = 5$ and $U = 3$. We also evaluate the performance with $L' < L$, that is, the reduced number of classes for computational efficiency. In our simulations, the number of training samples is set by 4.9×10^4 , i.e., $D = 4.9 \times 10^4$, which is the number of rows in the training data matrix \mathbf{T} . For a k -NN algorithm, we set $k = \frac{D}{100}$ and use a Euclidian distance metric for the best classification accuracy of the antenna allocation system. For ease of presentation, we evaluate the performance by limiting the number of users, U , to certain values (e.g., $U = 3$). However, we can adopt any U in our system by generating multiple machine learning models offline according to various U s and then choosing one trained model for a given U .

4.1. Sum-Rate Performance

The sum-rate performance of SVM designed in the previous section is compared to the other schemes. In Figure 5, we illustrate the cumulative density function (CDF) of the sum rate when $N_t = 5$ and $U = 3$. It can be seen that the SVM classifier provides the closest sum-rate performance to that of OPT. The k -NN classifier achieves superior sum-rate performance to that of RAND, yet it is highly inferior to the proposed SVM.

In Figures 6 and 7, we illustrate the sum rate over SNR for different numbers of users $U \in \{1, 3\}$, when $N_t = 5$. For $U = 1$ corresponding to single-user communications with no interference, the sum rate curves of SVM and k -NN almost coincide with that of OPT due to the excellent classification accuracy of our learning system. On the other hand, for $U = 3$, we observe that SVM outperforms k -NN, as shown in Figure 5. The sum-rate performance of all methods gets saturated in a high SNR regime because our communication system is interference-limited.

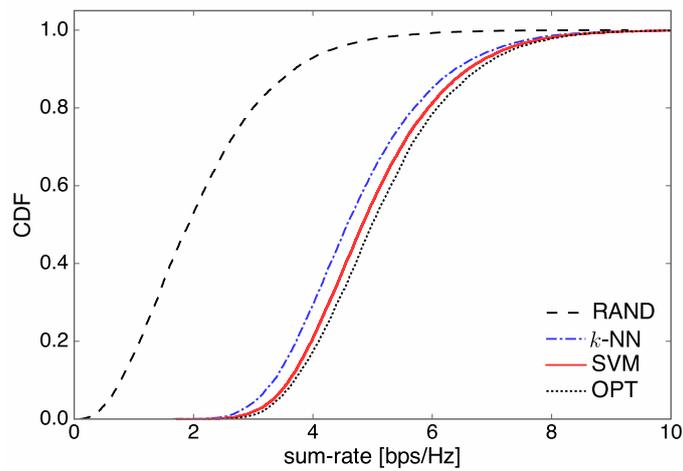


Figure 5. Empirical CDF of sum rate when $N_t = 5$ and $U = 3$.

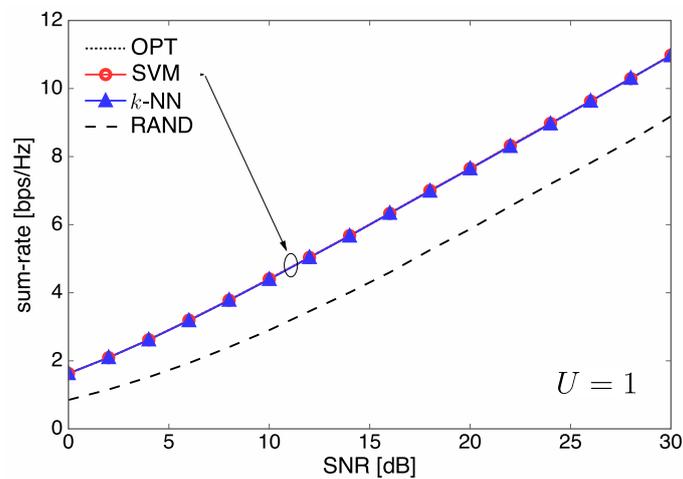


Figure 6. Sum rate over signal-to-noise ratio (SNR) performance for different system configurations when $N_t = 5$ and $U = 1$.

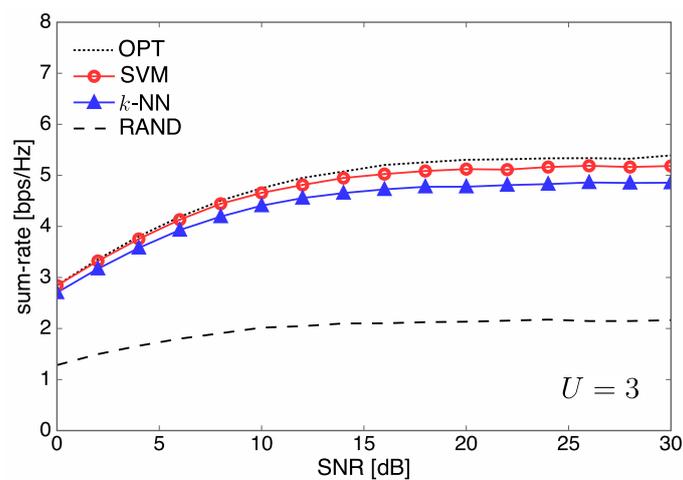


Figure 7. Sum rate over SNR performance for different system configurations when $N_t = 5$ and $U = 3$.

In Figure 8, we plot the sum rate over N_t from 5 to 10 when $U = 3$. The number of classes L is set according to the value of N_t , e.g., $L = 10!/7! = 720$ for $N_t = 10$. We observe that the sum rate of SVM and OPT increases with N_t with the help of the multi-antenna selection diversity gain. Interestingly,

the sum-rate performance of k -NN tends to saturate with increasing N_t due to the limited number of training samples. More specifically, while there are 720 classes, the number of training samples is not sufficient to guarantee classification performance in the nearest neighbor search. Thus, SVM is shown to be more robust to various system configurations scalable with N_t .

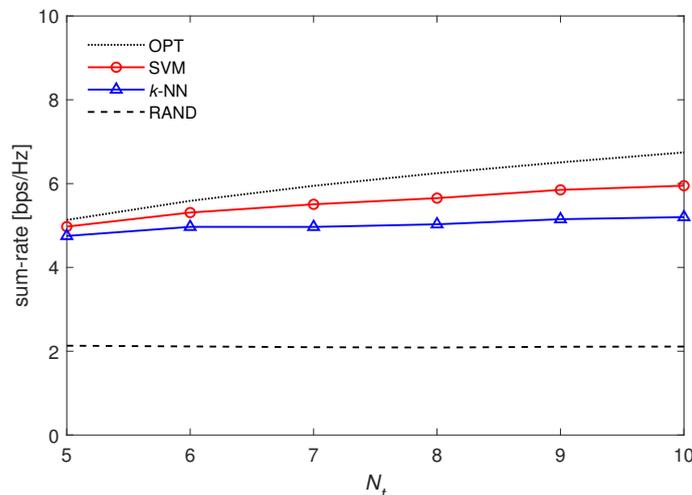


Figure 8. Sum rate over N_t when $U = 3$.

In Figure 9, we compare the sum rate over the number of classes L' when $N_t = 5$, $U = 3$ and $L' \in \{1, 2, 4, 8, 16, 32, 60\}$. It is observed that the sum rates of SVM and k -NN coincide with that of RAND when $L' = 1$, and they increase monotonically with L' .

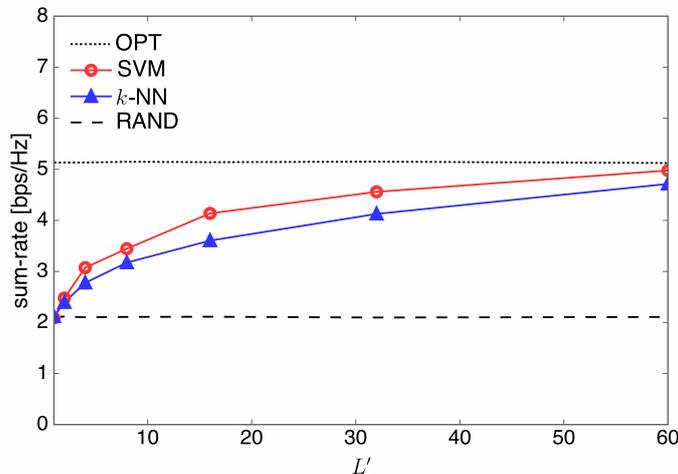


Figure 9. Sum rate over L' when $N_t = 5$ and $U = 3$.

To demonstrate our SVM-based approach in a massive multi-antenna setting, in Figure 10 we plot the sum rate over N_t from 20 to 100 when $U = 2$ and $L' = 60$. It is observed that, unlike the results in Figure 10, the sum rate of SVM is reduced with increasing N_t . Such a degradation occurs because the number of classes, L' , is over-reduced for large N_t . To overcome this problem, L' needs to scale according to the size of N_t , which is not employed in our study, however, since such a scaling of L' should be accompanied by a much larger training dataset, which may cause a memory overflow, in order to guarantee a sum-rate performance comparable to that of OPT. Using a sophisticated offline training method to appropriately adjust the number of classes without any memory overflow remains a goal for future work. However, it can be seen that SVM still offers substantial gains in terms of sum rate compared to RAND.

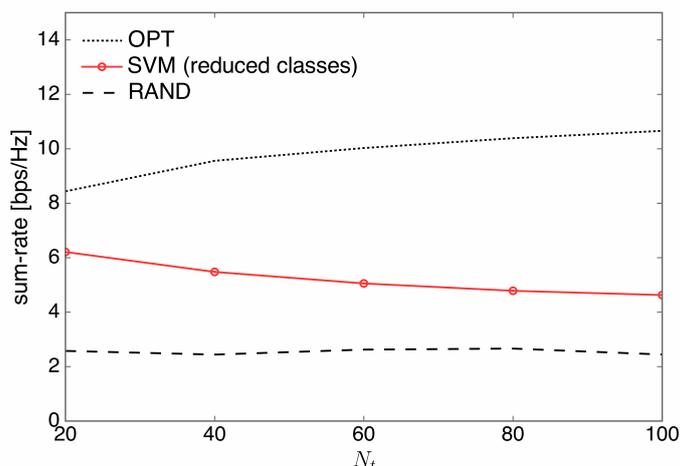


Figure 10. Sum rate over N_t when $L' = 60$ and $U = 2$.

4.2. Complexity Analysis and Runtime Evaluation

Selection complexities are compared in Table 2. The complexity of the optimization-based algorithm is discussed in Section 2.2, and the complexity of random antenna selection is $\mathcal{O}(1)$ because the only task is to generate a random integer. For the machine learning-based algorithms, selection complexity is defined as the online prediction complexity, excluding the training complexity since training can be completed offline with more powerful computing resources before the communication phase. In the k -NN algorithm, the complexity of $\mathcal{O}(DN_tU)$ and $\mathcal{O}(Dk)$ comes from the computing distances to all training samples and finding the k nearest neighbors, respectively [58]. In multiclass SVM algorithms, for the OVO strategy, because $M = \frac{L(L-1)}{2}$ binary SVM learners are employed for the multiclass prediction, complexity is given by $\mathcal{O}(L^2N_tU)$; while for the OVA strategy, L binary SVM learners are required, and thus complexity is $\mathcal{O}(L'N_tU)$ [57]. Note that the complexity of the learning-based methods mainly depends on the number of classes, and the number of classes can be reduced from L to L' , as discussed in Section 3.1. From Table 2, it can be seen that the selection complexity of the machine learning-based algorithms (k -NN and SVM) is polynomial on N_t and U , which is lower than that of the optimization-based algorithm using exhaustive search among all potential antenna allocation schemes.

Now, the runtime complexity (in seconds) of the online antenna selection is evaluated under the various simulation environments. Figure 11 shows runtime complexity over L' when $N_t = 5$, $U = 3$, and $L' \in \{1, 2, 4, 8, 16, 32, 60\}$. It can be seen that the complexity of k -NN is much higher than that of OPT when $L' \geq 4$. Since the complexity of k -NN is proportional to D , as shown in Table 2, the complexity of k -NN can be reduced at the cost of a degraded sum rate. On the other hand, the complexity of SVM is much lower than that of OPT; therefore, SVM is favorable for our antenna selection system.

Table 2. Performance and allocation complexity of the algorithms. OVO: one-vs-one strategy; OVA: one-vs-all strategy

Algorithm	OPT	RAND	k -NN	SVM
Sum-rate performance	Best	Worst	Close to SVM	Second-best
Allocation complexity	$\mathcal{O}\left(\frac{U^2N_t!}{(N_t-U)!}\right)$	$\mathcal{O}(1)$	$\mathcal{O}(DN_tU + Dk)$	$\mathcal{O}(L^2N_tU)$ for OVO and $\mathcal{O}(L'N_tU)$ for OVA

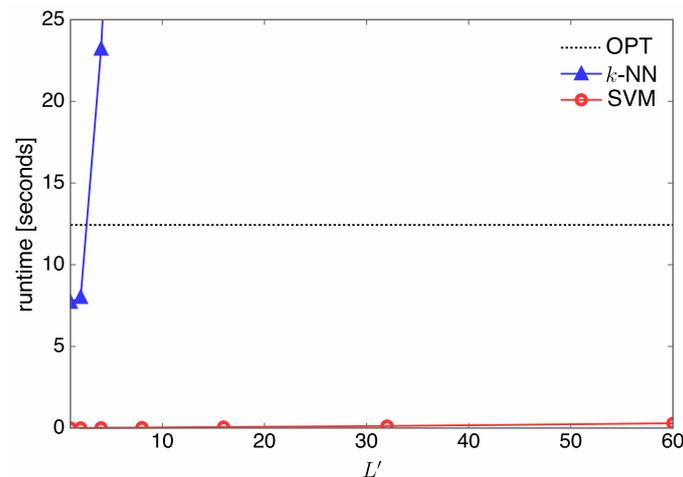


Figure 11. Runtime over L' when $N_t = 5$ and $U = 3$.

In Figure 12, we plot the runtime complexity over N_t from 5 to 12 when $U = 3$. We observe that the complexity of SVM is further reduced and remarkably lower than that of OPT when $L' = 32$. More specifically, the complexity of SVM tends to increase slowly on a linear scale with N_t , whereas that of OPT dramatically increases as N_t increases (refer to Table 2). However, the complexity of k -NN is still much greater than that of OPT.

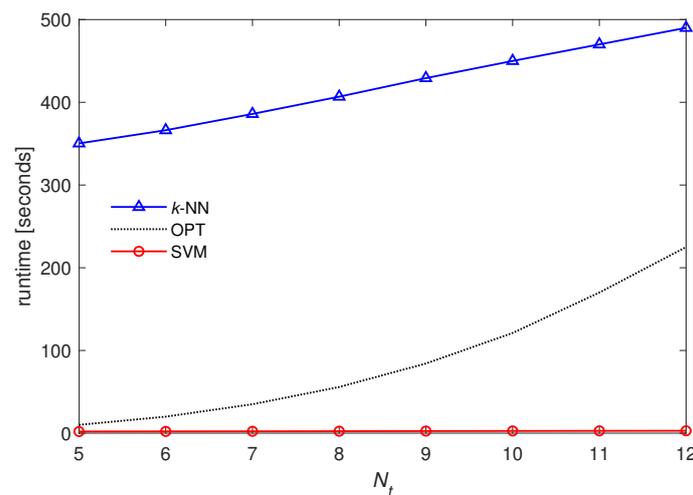


Figure 12. Runtime over N_t when $U = 3$ and $L' = 32$.

5. Concluding Remarks and Future Work

In this paper, we introduced a new framework for applying multiclass classification to an antenna allocation system with multiple antennas in multiuser downlink communications, under the assumption of channel amplitude information at the transmitter. The proposed antenna allocation system based on an SVM multiclass classifier was numerically evaluated and verified based on sum-rate performance and computational complexity. The following main results were obtained: (i) If the number of classes L' is suitably established, then the sum-rate performance of SVM is comparable to that of the optimization-driven method and significantly reduces computational complexity in the online antenna section; (ii) the classification performance of k -NN is inferior but still comparable to that of SVM; and (iii) for a given L' , the runtime complexity of the SVM classifier increases linearly with the number of antennas, which implies that the designed learning-based approach using SVM is appropriate, especially for large-scale antenna systems. Suggestions for future research in this area include (i) developing a variety of learning systems by precisely designing training data along

with more channel information (e.g., channel phase information) in interference-limited multiuser communications, (ii) supporting multi-antenna users, and (iii) developing an online learning algorithm to track channels with time-varying statistics.

Author Contributions: Conceptualization, J.J. and W.-Y.S.; methodology, J.J.; software, H.L.; validation, H.L.; investigation, H.L., W.-Y.S. and J.J.; writing—original draft preparation, H.L.; writing—review and editing, W.-Y.S. and J.J.; visualization, J.J.; supervision, J.J.; project administration, J.J.; funding acquisition, J.J.

Funding: This research was supported by the National Research Foundation of Korea (NRF) grant (2018R1A4A1023826) funded by the Korean government (MSIT).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Er, M.J.; Zhou, Y. *Theory and Novel Applications of Machine Learning*; InTech.: London, UK, 2009.
2. Michalski, R.S.; Carbonell, J.G.; Mitchell, T.M. *Machine Learning: An Artificial Intelligence Approach*. Springer Science & Business Media: Berlin, Germany, 2013.
3. Zhang, Z.; Chong, E.K.P.; Pezeshki, A.; Moran, W.; Howard, S.D. Learning in hierarchical social networks. *IEEE J. Sel. Topics Signal Process.* **2013**, *7*, 305–317. [[CrossRef](#)]
4. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2016.
5. Xu, J.; Song, L.; Xu, J.Y.; Pottie, G.J.; Van der Schaar, M. Personalized active learning for activity classification using wireless wearable sensors. *IEEE J. Sel. Topics Signal Process.* **2016**, *10*, 865–876. [[CrossRef](#)]
6. Chen, X.; Kopsaftopoulos, F.; Wu, Q.; Ren, H.; Chang, F.K. Flight state identification of a self-sensing wing via an improved feature selection method and machine learning approaches. *Sensors* **2018**, *18*, 1379. [[CrossRef](#)]
7. Liu, D.; Liu, X.; Wu, Y. Depth reconstruction from single images using a convolutional neural network and a condition random field model. *Sensors* **2018**, *18*, 1318. [[CrossRef](#)] [[PubMed](#)]
8. Hosseinyalamdary, S. Deep Kalman filter: Simultaneous multi-sensor integration and modelling; A GNSS/IMU case study. *Sensors* **2018**, *18*, 1316. [[CrossRef](#)] [[PubMed](#)]
9. Woo, W.L.; Gao, B.; Bouridane, A.; Ling, B.W.; Chin, C.S. Unsupervised learning for monaural source separation using maximization-minimization algorithm with time-frequency deconvolution. *Sensors* **2018**, *18*, 1371. [[CrossRef](#)]
10. Yang, G.; Yang, J.; Sheng, W.; Junior, F.; Li, S. Convolutional neural network-based embarrassing situation detection under camera for social robot in smart homes. *Sensors* **2018**, *18*, 1530. [[CrossRef](#)]
11. Coutinho, M.; de Oliveira Albuquerque, R.; Borges, F.; García Villalba, L.J.; Kim, T.H. Learning perfectly secure cryptography to protect communications with adversarial neural cryptography. *Sensors* **2018**, *18*, 1306. [[CrossRef](#)] [[PubMed](#)]
12. Heath, R.W.; Sandhu, S.; Paulraj, A. Antenna selection for spatial multiplexing systems with linear receivers. *IEEE Commun. Lett.* **2001**, *5*, 142–144. [[CrossRef](#)]
13. Sanayei, S.; Nosratinia, A. Antenna selection in MIMO systems. *IEEE Commun. Mag.* **2004**, *42*, 68–73. [[CrossRef](#)]
14. Molisch, A.F.; Win, M.Z.; Choi, Y.-S.; Winters, J.H. Capacity of MIMO systems with antenna selection. *IEEE Trans. Wirel. Commun.* **2005**, *4*, 1759–1772. [[CrossRef](#)]
15. Chen, R.; Heath, R.W.; Andrews, J.G. Transmit selection diversity for unitary precoded multiuser spatial multiplexing systems with linear receivers. *IEEE Trans. Signal Process.* **2007**, *55*, 1159–1171. [[CrossRef](#)]
16. Samek, W.; Stanczak, S.; Wiegand, T. The convergence of machine learning and communications. *arXiv* **2017**, arXiv: 1708.08299.
17. Jiang, C.; Zhang, H.; Ren, Y.; Han, Z.; Chen, K.-C.; Hanzo, L. Machine learning paradigms for next-generation wireless networks. *IEEE Wireless Commun.* **2017**, *24*, 98–105. [[CrossRef](#)]
18. Bi, S.; Zhang, R.; Ding, Z.; Cui, S. Wireless communications in the era of big data. *IEEE Commun. Mag.* **2015**, *53*, 190–199. [[CrossRef](#)]
19. Nguyen, T.T.; Armitage, G. A survey of techniques for internet traffic classification using machine learning. *IEEE Commun. Surveys Tuts.* **2008**, *10*, 56–76. [[CrossRef](#)]
20. Molisch, A.F.; Win, M.Z. MIMO systems with antenna selection. *IEEE Microw. Mag.* **2004**, *5*, 46–56. [[CrossRef](#)]

21. Gorokhov, A.; Gore, D.A.; Paulraj, A.J. Receive antenna selection for MIMO spatial multiplexing: Theory and algorithms. *IEEE Trans. Signal Process.* **2003**, *51*, 2796–2807. [[CrossRef](#)]
22. Choi, Y.; Molisch, A.; Win, M.; Winters, J. Fast antenna selection algorithms for MIMO systems. In Proceedings of the IEEE Vehicular Technology Conference, Orlando, FL, USA, 6–9 October 2003; pp. 1733–1737.
23. Lin, P.-H.; Tsai, S.-H. Performance analysis and algorithm designs for transmit antenna selection in linearly precoded multiuser MIMO systems. *IEEE Trans. Veh. Technol.* **2012**, *61*, 1698–1708. [[CrossRef](#)]
24. Patcharamaneepakorn, P.; Doufexi, A.; Armour, S. Reduced complexity joint user and receive antenna selection algorithms for SLNR-based precoding in MU-MIMO systems. In Proceedings of the 2012 IEEE 75th Vehicular Technology Conference (VTC Spring), Yokohama, Japan, 6–9 May 2012; pp. 1–5.
25. Jeong, M.-W.; Ban, T.-W.; Jung, B.C. User and antenna joint selection in multi-user large-scale MIMO downlink networks. *IEICE Trans. Commun.* **2017**, *100*, 529–535. [[CrossRef](#)]
26. Chan, P.T.; Palaniswami, M.; Everitt, D. Neural network-based dynamic channel assignment for cellular mobile communication systems. *IEEE Trans. Veh. Technol.* **1994**, *43*, 279–288. [[CrossRef](#)]
27. Xia, M.; Owada, Y.; Inoue, M.; Harai, H. Optical and wireless hybrid access networks: Design and optimization. *J. Opt. Commun. Netw.* **2012**, *4*, 749–759. [[CrossRef](#)]
28. Joung, J.; Chia, Y.K.; Sun, S. Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 954–965. [[CrossRef](#)]
29. Joung, J. and Jung, B.C. Machine learning based blind decoding for space-time line code (STLC) systems. *IEEE Trans. Veh. Technol.* **2019**, *1*, 1–4.
30. Sharma, D.K.; Dhurandher, S.K.; Woungang, I.; Srivastava, R.K.; Mohanane, A.; Rodrigues, J.J.P.C. A machine learning-based protocol for efficient routing in opportunistic networks. *IEEE Syst. J.* **2018**, *12*, 2207–2213. [[CrossRef](#)]
31. Maghsudi, S.; Stańczak, S. Channel selection for network-assisted D2D communication via no-regret bandit learning with calibrated forecasting. *IEEE Trans. Wireless Commun.* **2015**, *14*, 1309–1322. [[CrossRef](#)]
32. Tekin, C.; Liu, M. Online learning methods for networking. *Found. Trends. Network.* **2015**, *8*, 281–409. [[CrossRef](#)]
33. Yun, S.; Caramanis, C. Reinforcement learning for link adaptation in MIMO-OFDM wireless systems. In Proceedings of the 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, Miami, FL, USA, 6–10 December 2010; pp. 1–5.
34. Iacoboiaea, O.-C.; Sayrac, B.; Jemaa, S.B.; Bianchi, P. SON coordination in heterogeneous networks: A reinforcement learning framework. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 5835–5847.
35. Aprem, A.; Murthy, C.R.; Mehta, N.B. Transmit power control policies for energy harvesting sensors with retransmissions. *IEEE J. Sel. Top. Signal Process.* **2013**, *7*, 895–906. [[CrossRef](#)]
36. Lundén, J.; Kulkarni, S.R.; Koivunen, V.; Poor, H.V. Multiagent reinforcement learning based spectrum sensing policies for cognitive radio networks. *IEEE J. Sel. Top. Signal Process.* **2013**, *7*, 858–868. [[CrossRef](#)]
37. Alnwaimi, G.; Vahid, S.; Moessner, K. Dynamic heterogeneous learning games for opportunistic access in LTE-based macro/femtocell deployments. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 2294–2308. [[CrossRef](#)]
38. Tsiligkaridis, T.; Romero, D. Accelerated reinforcement learning algorithms with nonparametric function approximation for opportunistic spectrum access. *arXiv* **2017**, arXiv:1706.04546.
39. Daniels, R.; Heath, R. An online learning framework for link adaptation in wireless networks. In Proceedings of the 2009 Information Theory and Applications Workshop, San Diego, CA, USA, 8–13 February 2009; pp. 138–140.
40. Yun, S.; Caramanis, C. Multiclass support vector machines for adaptation in MIMO-OFDM wireless systems. In Proceedings of the 2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 30 September–2 October 2009; pp. 1145–1152.
41. Rico-Alvarino, A.; Heath, R.W. Learning-based adaptive transmission for limited feedback multiuser MIMO-OFDM. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 3806–3820. [[CrossRef](#)]
42. Feng, C.; Liao, S. Scalable Gaussian kernel support vector machines with sublinear training time complexity. *Inf. Sci.* **2017**, *418*, 480–494. [[CrossRef](#)]
43. Yuan, Y.; Zhang, M.; Luo, P.; Ghassemlooy, Z.; Lang, L.; Wang, D.; Zhang, B.; Han, D. SVM-based detection in visible light communications. *Optik* **2017**, *151*, 55–64. [[CrossRef](#)]
44. Wang, J.-B.; Wang, J.; Wu, Y.; Wang, J.-Y.; Zhu, H.; Lin, M.; Wang, J. A machine learning framework for resource allocation assisted by cloud computing. *IEEE Netw.* **2018**, *32*, 144–151. [[CrossRef](#)]

45. Al Rahhal, M.M.; Bazi, Y.; Alhichri, H.; Alajlan, N.; Melgani, F.; Yager, R.R. Deep learning approach for active classification of electrocardiogram signals. *Inf. Sci.* **2016**, *345*, 340–354. [[CrossRef](#)]
46. O’Shea, T.; Hoydis, J. An introduction to deep learning for the physical layer. *IEEE Trans. Cogn. Commun. Netw.* **2017**, *3*, 563–575. [[CrossRef](#)]
47. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*. MIT Press: Cambridge, MA, USA, 2016.
48. Joung, J. Machine learning-based antenna selection in wireless communications. *IEEE Commun. Lett.* **2016**, *20*, 2241–2244. [[CrossRef](#)]
49. Asadi, A.; Mancuso, V. A survey on opportunistic scheduling in wireless communications. *IEEE Commun. Surv. Tuts.* **2013**, *15*, 1671–1688. [[CrossRef](#)]
50. Kurniawan, E.; Joung, J.; Sun, S. Limited feedback scheme for massive MIMO in mobile multiuser FDD systems. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; pp. 1710–1715.
51. Zhang, H.; Berg, A.C.; Maire, M.; Malik, J. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06), New York, NY, USA, 17–22 June 2006; pp. 2126–2136.
52. Bishop, C.M. *Pattern Recognition and Machine Learning*. Springer: New York, NY, USA, 2006.
53. Osseiran, A.; Boccardi, F.; Braun, V.; Kusume, K.; Marsch, P.; Maternia, M.; Queseth, O.; Schellmann, M.; Schotten, H.; Taoka, H.; et al. Scenarios for 5G mobile and wireless communications: The vision of the METIS project. *IEEE Commun. Mag.* **2014**, *52*, 26–35. [[CrossRef](#)]
54. Joung, J.; Sun, S. Two-step transmit antenna selection algorithms for massive MIMO. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 1–6.
55. Cariou, C.; Chehdi, K. Unsupervised nearest neighbors clustering with application to hyperspectral images. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 1105–1116. [[CrossRef](#)]
56. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
57. Escalera, S.; Pujol, O.; Radeva, P. On the decoding process in ternary error-correcting output codes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 120–134. [[CrossRef](#)]
58. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).