MACHINE LEARNING PAPER REVIEW

TRANSMIT ANTENNA SELECTION IN MIMO WIRETAP CHANNELS: A MACHINE LEARNING APPROACH.

Emmanuel Obeng Frimpong

15th April 2021

HANBAT NATIONAL UNIVERSITY

CONTENTS

- Motivation
- System Model
- Proposal

MOTIVATION

- Motivated by the benefits provided by multi-input multi-output (MIMO) techniques, such as high reliability and high data rate, physical layer security in MIMO wiretap channels has recently attracted significant research attention.
- MIMO techniques, such as beamforming, artificial noise, and transmit antenna selection, in physical layer security can also be characterized as a classification problem and/or a decision-making problem.
- Paper leverages machine learning to enhance physical layer security in multi-input multi-output multi-antenna-eavesdropper wiretap channels.

SYSTEM MODEL

- Paper assumes that the source adopts TAS as the transmission strategy.
- Characterizes the problem of TAS as a multi-class classification problem.
- Then we propose support vector machine (SVM) and naive-Bayes (NB) schemes to select the optimal antenna that maximizes the secrecy performance of the considered system.

SYSTEM MODEL – CONVENTIONAL TAS

- Consider a wiretap channel, where a source transmits to a legitimate receiver in the presence of an eavesdropper. The source, the legitimate receiver, and the eavesdropper are equipped with Ns, Nr and Ne antennas, respectively.
- We also assume that **H** is known at the source, while **G** can be either known or not known at the source.
- **H is** the main channel between the source and the legitimate receiver.
- **G is** the channel between the source and the eavesdropper.

SYSTEM MODEL – CONVENTIONAL TAS

- In the conventional TAS scheme, the source selects one of its Ns antennas that maximizes the secrecy performance to transmit the confidential message.
- Then, the legitimate receiver and the eavesdropper adopt maximal ratio combining as the receiving strategy.

MACHINE LEARNING BASED TAS

- *** INTUITION

- We extract features from the CSIs, we then apply two different machine learning approaches, namely, SVM and NB, to construct the classification model and predict the class label that the current channel belongs to.

- We note that the belonged class represents an ideal antenna index to select that may optimize the secrecy performance of the current channel.

PREPARATION OF LEARNING

- Generate the feature vector for each training CSI example

Absolute values from each element of H and G as the feature vector containing complex-valued elements, feature vectors contain real-valued elements extracted from the training CSI matrices. Specifically, we first choose the absolute value of each element of $H_m \; and \; G_m$ as the element of the feature vector. Then we normalize the feature vectors in order to avoid bias in the learning process.

FEATURE GENERATION DETAILED

Step 1: Generate a $1 \times N$ real vector $\mathbf{d}^m = [d_1^m, d_2^m, \cdots, d_N^m]$, where N equals to $N_s \times (N_r + N_e)$ for the full case and $N_s \times N_r$ for the partial case. For the full CSI case, $\mathbf{d}^m$ is expressed as

$$\mathbf{d}^m = \Big[|h_{1,1}^m|, \ldots, |h_{1,N_s}^m|, |h_{2,1}^m|, \ldots, |h_{N_r,N_s}^m|, \ldots, |g_{1,1}^m|, \ldots, |g_{1,N_s}^m|, |g_{2,1}^m|, |g_{2,2}^m|, |g_{N_e,N_s}^m|\Big], \quad (3)$$

and for the partial CSI case, $\mathbf{d}^m$ is expressed as

$$\mathbf{d}^m = \Big[|h_{1,1}^m|, \ldots, |h_{1,N_s}^m|, |h_{2,1}^m|, |h_{2,2}^m|, \ldots, |h_{N_r,N_s}^m|\Big], \quad (4)$$

where $h_{i,j}^m$ and $g_{i,j}^m$ denote the $(i,j)th$ element of $\mathbf{H}^m$ and $\mathbf{G}^m$, respectively.

Step 2: Repeat **Step 1** for all M training CSI examples and generate M feature vectors, i.e., $\mathbf{d}^1, \mathbf{d}^2, \cdots, \mathbf{d}^M$.

Step 3: Normalize $\mathbf{d}^m$ and generate the normalized feature vector $\mathbf{t}^m \in \mathbb{R}^{1\times N}$, for $m \in \{1, 2, \ldots, M\}$. The n-th element of $\mathbf{t}^m$ can be expressed as

$$t_n^m = \big(d_n^m - \mathbb{E}[\mathbf{d}^m]\big)/\big(\max(\mathbf{d}^m) - \min(\mathbf{d}^m)\big), \quad (5)$$

where d_n^m is the n-th element of $\mathbf{d}^m$.

**from paper*

PREPARATION OF LEARNING

- *KPI DESIGN:*

- KPI is the metric to classify the training CSI examples. Aiming at maximizing the secrecy performance, we choose the achievable secrecy rate as the KPI for the full CSI case. As for the partial CSI case, we adopt the achievable rate of the main channel C_b as the KPI.

- *CLASSIFICATION OF TRAINING CSI EXAMPLES:*

- In order to determine the class label of one training CSI example, we first calculate the KPI for each antenna. Then we choose the class label of this training CSI example as the index of the antenna that achieves the maximum KPI.

SVM-Based Scheme

- Paper constructs an SVM model of one-against-the-rest using the learning parameter.

$$\boldsymbol{w}_l = \underset{\boldsymbol{w}_l}{\operatorname{argmin}}\, C \sum_{m=1}^{\mathrm{M}} \left[b_l[m] \operatorname{cost}_1\left(\boldsymbol{w}_l^T \boldsymbol{f}\left(\mathbf{t}^m\right)\right) + \left(1 - b_l[m]\right) \times \operatorname{cost}_0\left(\boldsymbol{w}_l^T \boldsymbol{f}\left(\mathbf{t}^m\right)\right) \right] + \|\boldsymbol{w}_l\|_2^2/2,$$

- For a new CSI example, we generate a normalized feature vector t using the feature vector generation discussed. The class label of current channel can be determined by using t to replace t_m.
- Then the class label of current channel, (the antenna that should be selected) is the one that achieves the largest $w_l^T f(t)$ among all classes.

NB-BASED SCHEME

- The NB classification model is constructed by calculating the probability distribution for each element of the normalized feature vector for all label classes.

$$\Pr(c = l|\mathbf{t}) = \frac{\Pr(\mathbf{t}|c = l)\Pr(c = l)}{\Pr(\mathbf{t})} = \frac{\prod_{n=1}^{N}\Pr(\mathrm{t}_n|c = l)\Pr(c = l)}{\Pr(\mathbf{t})},$$

- The problem of selecting the optimal antenna can be reduced to the problem of selecting the class label that achieves the maximal probability of the occurrence of feature vector t.

$$l^* = \underset{l \in \{1,2\ldots,N_s\}}{\operatorname{argmax}} \prod_{n=1}^{N} \Pr(\mathrm{t}_n|c = l).$$

NB-BASED SCHEME

1) Find all the feature-label pairs $\{\mathbf{t}^m, \mathrm{c}^m\}$ that satisfies $\mathrm{c}^m = l$ and construct a new set $\mathbb{L} = \{\mathbf{t}^m | \mathrm{c}^m = l, m \in \{1, 2, \ldots, M\}\}$;
2) Use all t_n^m of $\mathbf{t}^{\mathrm{m}} \in \mathbb{L}$ to calculate the probability distribution of $\Pr(\mathrm{t}_n | c = l)$ for $n \in \{1, 2, \ldots N\}$;
3) Repeat 1) and 2) for all $l \in \{1, 2, \ldots, N_s\}$.

We note that the NB-based learning model is ready for TAS after the probability set, given by $\mathbb{P} = \{\Pr(\mathrm{t}_n | c = l) | n \in \{1, 2, \ldots N\}, l \in \{1, 2, \ldots N_s\}\}$, is obtained.

**from paper*

NUMERICAL RESULTS

- Maximum Achievable Secrecy Rate

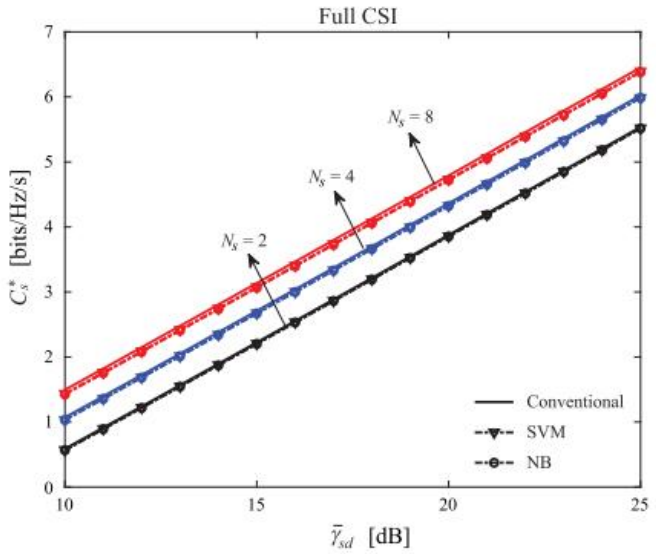


Fig. 1. C_s^* versus $\overline{\gamma}_{sd}$ for different transmission schemes and values of N_s with $\overline{\gamma}_{se} = 10$ dB.

NUMERICAL RESULTS

- Minimum SOP

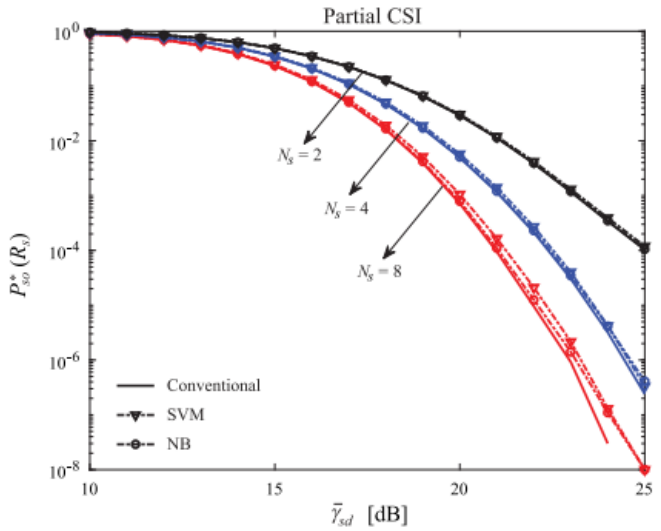


Fig. 2. $P^{*}_{so}(R_s)$ versus $\overline{\gamma}_{sd}$ for different transmission schemes and values of N_s with $\overline{\gamma}_{se} = 10$ dB and $R_s = 2$ bits/Hz/s.

NUMERICAL RESULTS

- Performance of the machine learning based TAS scheme using the normalized mean square error.

$$\mathrm{NMSE} = 10\log_{10}\frac{||\mathbf{s}_{\mathrm{con}} - \mathbf{s}_{\mathrm{ML}}||^2}{||\mathbf{s}_{\mathrm{con}}||^2},$$

TABLE I
NMSEs OF DIFFERENT TAS SCHEMES

Scheme	Full CSI		Partial CSI	
	SVM	NB	SVM	NB
$N_s = 2$	−45.14 dB	−43.61 dB	−47.74 dB	−62.69 dB
$N_s = 4$	−39.65 dB	−40.24 dB	−39.48 dB	−57.69 dB
$N_s = 8$	−37.18 dB	−35.48 dB	−34.58 dB	−48.87 dB

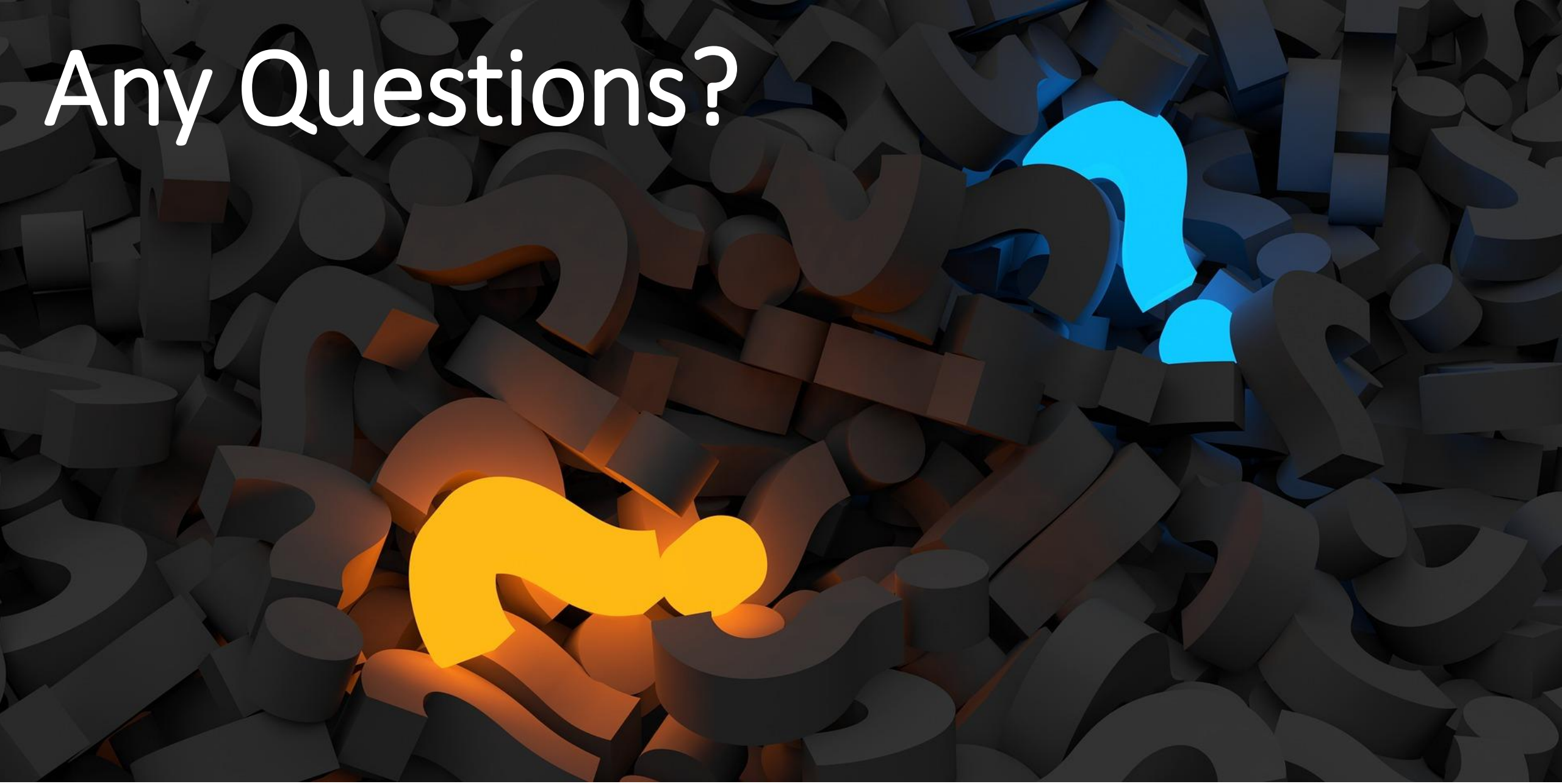
NUMERICAL RESULTS

- Complexities and Overheads

COMPLEXITIES AND OVERHEADS OF DIFFERENT TAS SCHEMES

Scheme	SVM	NB	Conventional
Selection complexity	$\mathcal{O}\left(N^2\right)$	$\mathcal{O}\left(\lvert\mathcal{L}\rvert N + \lvert\mathcal{L}\rvert \log\left(\lvert\mathcal{L}\rvert\right)\right)$	$\mathcal{O}\left(N + \lvert\mathcal{L}\rvert \log\left(\lvert\mathcal{L}\rvert\right)\right)$
Feedback overhead	N real values		N complex values

Any Questions?



frimpongemmanuelobeng@gmail.com