Multi-Receiver Modulation Classification for Non-Cooperative Scenarios

Garrett Vanhoy, Hamed Asadi, Haris Volos, Tamal Bose
Department of Electrical and Computer Engineering
The University of Arizona
Tucson, AZ 85721-0104
{gvanhoy, hasadi, hvolos, tbose}@arizona.edu

***Abstract*—Modulation Classification (MC) is a difficult task that can increase awareness in Cognitive Radio (CR) applications. Much of the research in MC has been for single antenna and single user scenarios. For multiple users, multiple receivers must be used to first separate the incoming signals before MC can be done. For non-cooperative communications blind source separation (BSS) techniques can readily separate a linear mixture of signals, but it is not clear which technique is best suited for MC. In this work, we compare three BSS algorithms as candidates for multi-receiver MC while examining both single and multiple user MC. In our simulations, the fastICA algorithm achieves the best performance of the three [1]. At 0 dB SNR for a single user, the fastICA algorithm acheives 96% correct classification and 92% for multiple users with three receivers. This work also reveals that the combination of phase correction, fastICA, and support vector machines (SVMs) can achieve near-optimal performance.**

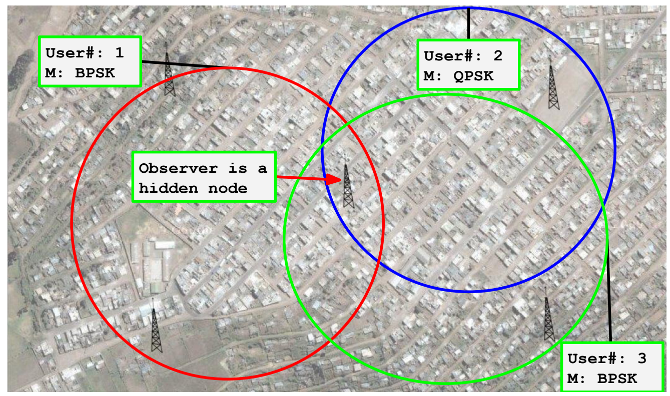

Fig. 1. Multi-User Scenario with the Observer as a Hidden Node

I. Introduction

Cognitive Radio (CR) is a technology that enables a radio to make intelligent decisions by exploiting knowledge about the radio environment. The more a radio is aware of its environment, the greater its ability to adapt to it using increasingly complex behaviors. The lowest level of awareness entails knowing whether or not a signal is present in a region of interest and this has been the subject of much research primarily for applications in dynamic spectrum access (DSA). A higher level of awareness could entail knowing which modulation is being transmitted. The task of determining the modulation has been named modulation classification (MC), recognition, or identification and has a variety of applications in the military and commercial domains [2]. Sometimes this is called *automatic* modulation classification in the sense that the classification decision is made automatically by a computer rather than a human being. In this paper we will focus on MC applications involving non-cooperative communications. In short non-cooperative communication is when the signal of interest is coming from a transmitter that does not intend for its data to be interpreted by the observing radio. This is in general a more challenging scheme than cooperative communications in which some properties of the incoming signal may be known *a-priori*, such as coding scheme, that can be exploited for better performance in MC [3].

Much of the research in MC has been focused on scenarios with a single observed signal and a single receiving antenna [4]. Some research has been done using multiple receiving antennas to increase classification performance, classify multiple signals, and implement distributed classification techniques [5]–[7]. The methods used in each application range from likelihood-based methods to a combination of Independent Component Analysis (ICA) and feature-based methods.

In this work MC algorithms are compared for both the single and multiple user scenarios. The blind source separation (BSS) techniques considered in this work include the (Joint Approximate Diagonal Eigenmatrices) JADE, Gönen, and fastICA algorithms [1], [8], [9]. Although each algorithm can perform BSS, a comparison of their performance in the context of MC does not exist to the best of our knowledge. Additionally, each BSS algorithm has limitations with regards to which signals it can separate. For example, in a single user scenario many BSS techniques can be leveraged to combine copies of the signal to increase the effectiveness of the classifier. However, some BSS techniques struggle to do this and thus their performances are compared to merely combining the classifications of independent receivers. The results show that even a combination of classifications can achieve reasonable performance and in some cases, better performance than BSS techniques. For multiple user scenarios, the system can be treated as a generic MIMO system with noise. However, it is important to note that in some scenarios, the observing radio will be the hidden node between several independent transmissions as in Figure 1. In this situation, transmitters may be a part of a larger system that prevents interference between each node. However, the transmitters are not aware of the observing radio and thus receives more than one signal simultaneously. In this case frequency offsets, timing offsets, and symbol rates can vary between overlapping transmissions. A treatment of separating this class of signals is left as an expansion on this work.

In Section II we outline our models and assumptions. Then, each stage of the proposed MC algorithm is explained in Section III. In Section IV, we compare results for different algorithms in both single and multi-user scenarios. Lastly, we conclude in Section V.

II. MODELS

The received noiseless digital baseband signal is commonly represented by (1) from [4]:

$$x_i(t) = \alpha_i e^{j2\pi\Delta_i f t} e^{j\theta_i} \sum_{k=1}^{K} e^{\phi_k} s_k g(t-(k-1)T - \epsilon T) \quad (1)$$

with Δ_i, θ_i, ϵ, and α_i, being the carrier frequency offset, the time-invariant carrier phase, and timing offset with respect to the receiver's reference clock, and the signal amplitude of the i^{th} signal respectively. Phase jitter is represented by ϕ_k for the k^{th} equi-probable complex data symbol s_k with symbol period T for the modulation of order M. The channel response to the transmitted pulse shape p_{TX} is the convolution between $h(t)$ and p_{TX} and is denoted by $g(t)$. Each signal impinges upon a uniform linear array (ULA) of isotropic elements. The signals arriving at each of the elements at separate times creates a change in phase of the received signal according to:

$$r_p = e^{\frac{j2\pi dp \sin(\Theta_i)}{\lambda}} x_i \quad (2)$$

With r_p representing the received signal on the p^{th} element, Θ_i being the angle of the incoming signal with respect to the array normal, d being the distance separating each element, and λ being the wavelength of the incoming narrowband signal. There are a total of P receiving elements and N incoming signals. For multiple incoming signals this can be extended to:

$$\begin{bmatrix} r_1 \\ \vdots \\ r_P \end{bmatrix} = \begin{bmatrix} e^{\frac{j2\pi d(0)\sin(\theta_1)}{\lambda}} & \dots & e^{\frac{j2\pi d(0)\sin(\theta_N)}{\lambda}} \\ \vdots & \ddots & \vdots \\ e^{\frac{j2\pi dP\sin(\theta_1)}{\lambda}} & \dots & e^{\frac{j2\pi dP\sin(\theta_N)}{\lambda}} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \quad (3)$$

The matrix factor of $\boldsymbol{x}$, named $\boldsymbol{A}$, is sometimes called the *steering matrix*. With the addition of additive white Gaussian noise, this is expressed:

$$\boldsymbol{r} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} \quad (4)$$

III. MODULATION CLASSIFICATION

Modulation classification can be separated into three stages that each have its own challenges. First, the preprocessing stage aims to accurately recover the transmitted signal and apply appropriate transformations. This is done to maximize the effectiveness of the second stage, the feature extraction stage, in which a set of features are extracted from the signal. Lastly the classification stage implements a decision structure to best differentiate the incoming signals based on the extracted features.

A. Preprocessing

For digital modulations, the preprocessing stage includes common procedures for synchronization such as carrier and timing recovery. It is often assumed that this procedure yields the ideally sampled digital symbols [10] while synchronization is left as a separate challenge. Additionally, the signal is normalized to have unit energy after the signals are separated. This simplifies the model in (1) to:

$$x_i(t) = e^{j\theta_i} \sum_{k=1}^{K} s_k \quad (5)$$

In many cases, synchronization cannot be done for low signal to noise ratio (SNR) without *a-priori* information. Therefore, we restrict our analyses to SNR above 0 dB.

1) Blind Source Separation: The JADE, Gönen, and fast ICA algorithms are investigated as candidates for separating the incoming signals. Both JADE and the Gönen algorithms were originally intended for use as beamforming algorithms. However, the JADE algorithm fails when sources have identical kurtoses, which is a statistic related to a fourth-order moments. This is generally not a problem for this application as signals on different antennas will have the addition of independent noise. The fastICA algorithm was developed for more general cases [1]. As such, general ICA approaches suffer from being unable to separate more than two Gaussian sources from each other. Each of these algorithms introduces an ambiguity in the phase of the separated signals and their respective orders.

2) Phase Correction: To maximize the effectiveness of the feature extraction stage, the phase ambiguity introduced through BSS is reduced by finding the phase that maximizes the average magnitude of the quadrature. Stated mathematically, let $s_k e^{\theta} = s_p$ with θ being determined from:

$$\arg\max_{\theta} \frac{1}{K} \sum_{k=0}^{K} |imag(s_k e^{j\theta})| \quad (6)$$

In this work, θ is found by a brute-force search algorithm which adds a considerable amount of unnecessary computational complexity. Through experimentation, it was found that checking 100 angles between 0 and π is sufficient. However, it would be reasonable to implement a search algorithm that can achieve a preferred level of stability in the phase correction. This method of correcting the phase contribution is unique and constitutes one of the key contributions of this work.

B. Feature Extraction

Feature extraction methods in MC can be largely separated into two types: likelihood-based and feature-based. Likelihood-based methods use likelihood ratios that can also be considered as extracted features. Generally likelihood-based methods can be considered optimal in the sense that they achieve the least probability of misclassification [11]. However, this comes at the cost of a computational complexity that eludes real-time implementation in many cases [4]. Likelihood-based methods come in three categories including the average likelihood ratio test (ALRT), the generalized likelihood ratio test (GLRT), and the hybrid likelihood ratio test (HLRT). The ALRT estimates the unknown parameters of a signal by treating them as random variables with known probability density functions (PDF). This is the most accurate of the likelihood-based approaches, but it is also the most computationally complex. For this reason, the ALRT is often used as a theoretical upper bound for the probability of correct classification for other methods. The GLRT uses maximum-likelihood estimates for the unknown quantities instead of treating them as random variables with known PDF. This is less computationally complex than the ALRT, but is less accurate than the ALRT and suffers from being unable to differentiate nested constellations such as 16-QAM and 64-QAM entirely [4]. These two algorithms can be combined to create the HLRT. Variants of the HLRT are the closest to real-time implementation with the Discrete Likelihood-Ratio Test (DLRT) being the only variant that has been implemented in a practical scenario [7].

A feature-based method extracts a set of descriptive values from the signal that differentiates each signal from each other. These features can include cumulants, statistics, Fourier Transform coefficients, Wavelet Transform coefficients, or a combination of them [12]. This approach is suboptimal in terms of probability of correct classification, but reduces the computational complexity to real-time applications. Finding the best set of features to accurately identify the modulation has been the subject of many papers. The most commonly used feature is fourth-order cumulants and other high-order cumulants which can readily distinguish linear digital modulation schemes from each other in low-SNR environments.

Two of the properties of fourth-order cumulants make them a desirable candidate as features. First, the cumulant of the sum of two independent distributions is the sum of the cumulants of the two distributions. Second, the cumulants of order higher than 3 for a Gaussian distribution is zero. Thus, the cumulants of a constellation with additive white Gaussian noise is ideally the cumulants of the constellation points without noise. The fourth-order cumulants $\hat{C}_{40}$ and $\hat{C}_{42}$ are considered for MC in this paper and are estimated using the following equations:

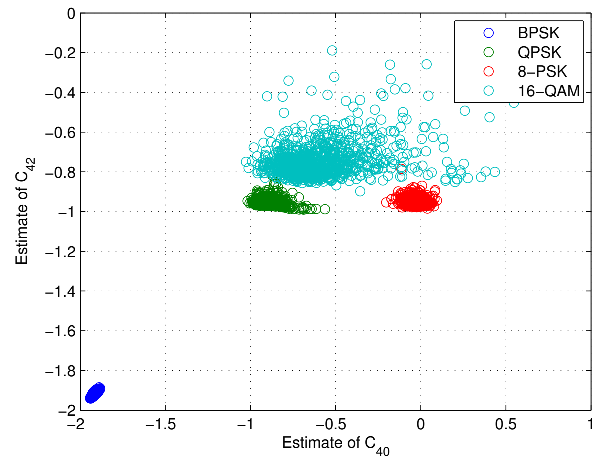

Fig. 2. Distribution of Cumulants

$$\hat{C}_{20} = \frac{1}{N}\sum_{n=1}^{N} y^2(n) \tag{7}$$

$$\hat{C}_{21} = \frac{1}{N}\sum_{n=1}^{N} |y(n)|^2 \tag{8}$$

$$\hat{C}_{40} = \frac{1}{N}\sum_{n=1}^{N} y^4(n) - 3\hat{C}_{20}^2 \tag{9}$$

$$\hat{C}_{42} = \frac{1}{N}\sum_{n=1}^{N} |y(n)|^4 - |\hat{C}_{20}|^2 - 2\hat{C}_{21}^2 \tag{10}$$

In Figure 2 the estimates for $|C_{40}|$ and $|C_{42}|$ for BPSK, QPSK, 8-PSK, and 16-QAM are plotted after a phase correction with 10 dB SNR. Notice that for each modulation scheme there are extreme points that appear as statistical outliers. These come from the tendency of the phase correction occasionally over or under-correct.

C. Classification

Since the phase correction portion given in Section III-A2 occasionally skews the cumulant estimates seen in Figure 2, the exact distribution of the cumulant estimates is unknown. Thus, using a type of pattern recognition for the classification stage is more appropriate than using a threshold. Support vector machines (SVMs) can be used for this type of data. Support vector machines were first introduced by Boser et al [13] in 1992. It has had successful applications in many fields that involve classification and fits into the broader study of supervised learning models. SVM is a learning algorithm that is widely used due to its ability to deal with high-dimensional data and efficiency in modeling diverse data. As a *supervised* learning algorithm an SVM is constructed offline by using a set of training data. It uses the training data to construct a hyperplane that optimally separates each class.

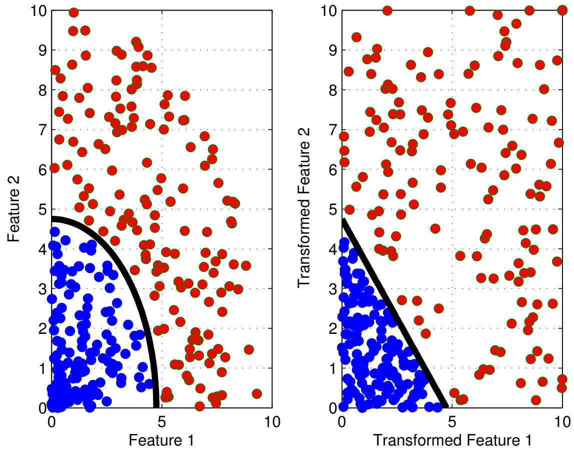

Fig. 3. (**Left**) Example Data Set with Separator (**Right**) Transformed Data Set with Optimal Hyperplane

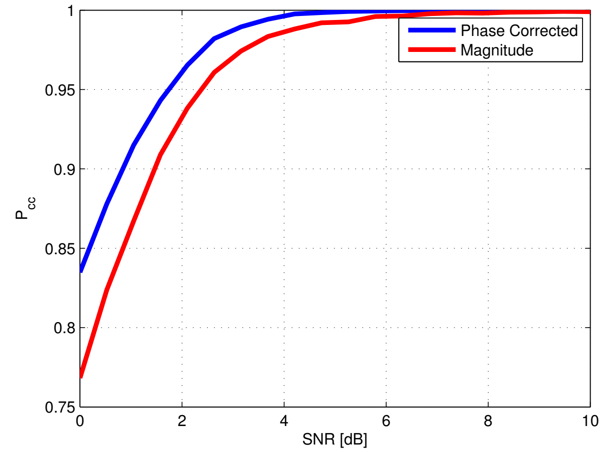

Fig. 4. Analysis of Phase Correction

This is sufficient only for linearly separable patterns, but can be extended to other patterns by transformations of the data.

The training process works by first mapping data to a feature space so that data points can be categorized. Then, a separator between the categories is found and the data are transformed in such a way that the separator could be drawn as a hyperplane. This hyperplane can be used to predict the group to which a new record should belong. As an example, consider the following Figure 3 in which the data points fall into two different categories. The two categories can be separated with a curve. After a transformation using a predefined kernel function, the boundary between the two categories can be defined by a hyperplane. In this paper we use LIBSVM [14] for our implementation using a linear kernel function.

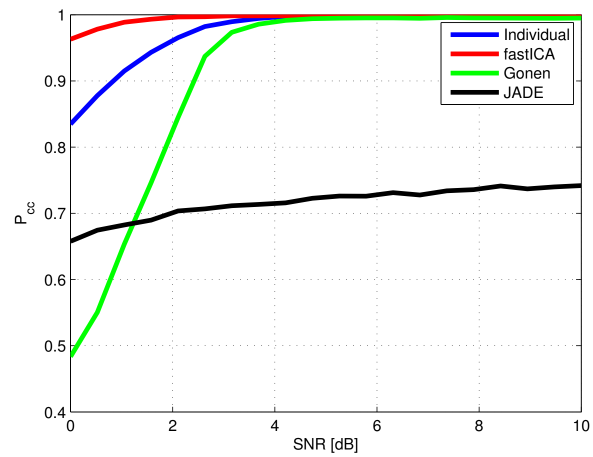

Fig. 5. Comparison of Single-User Methods

IV. Simulations

A. Single-User Scenarios

For a single user scenario, the same sequence of symbols is received by each antenna with additive white Gaussian noise (AWGN). The number of receiving antennas P is fixed at three and the number of symbols received T by the observing radio is 300. The set of possible modulations for each source is limited to BPSK, QPSK, 8-PSK, and 16-QAM. An example distribution of the real portion of the fourth-order cumulants for each of these modulations are depicted in Figure 2.

1) The Effect of Phase Correction: First, the effect of the phase correction outlined in Section III-A2 is examined. To examine this effect, the BSS methods are omitted in the preprocessing stage and estimates for each cumulant are generated from the signal received on each antenna. This creates a six-dimensional feature vector that the SVM uses to classify the incoming signal. The *magnitude* algorithm will contain estimates of the cumulant magnitudes and the *phase corrected* algorithm will contain the complex-valued estimates for the cumulants. Note that either of these methods can be used to resolve the ambiguity of the phase of the received signal.

The comparison of these two methods are seen in Figure 4. The *magnitude* algorithm performs slightly worse than the *phase corrected* algorithm. This suggests that for MC using the fourth-order cumulants, it is better to correct the phase of the constellation than to take the magnitude of the cumulant. Hence, it is used for the rest of the simulations in this paper.

2) ICA Algorithms for a Single User: Many ICA algorithms have limitations for the sets of signals it can resolve. This is highlighted especially in the single user case. Each of the three algorithms are compared along with the *phase corrected* algorithm in the previous section with the same simulation parameters.

It is evident from the results in Figure 5 that combining independent cumulant estimates performs better than using either the JADE or Gönen algorithms to increase the effective SNR. For example, at 0 SNR the phase corrected algorithm

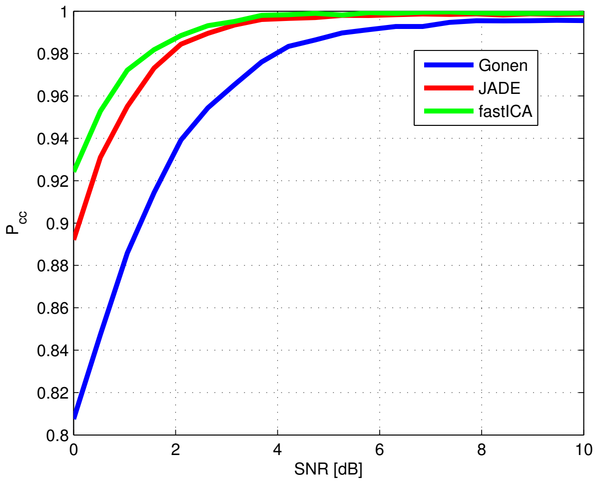

Fig. 6. Comparison of Multi-User Methods

achieves about 93 percent correct classification while the Gönen algorithm achieves only 49 percent. This is likely the case because BSS will not always succeed in increasing the effective SNR when the SNR of the signal is too low already. Thus, attempting to perform BSS may actually cause the estimates of the fourth-order cumulants to be worse.

B. Multi-User

For a multi-user scenario, data is sent independently by three transmitters. The data is simultaneously received by each antenna with additive white Gaussian noise (AWGN). Three receiving antennas receive at total of 300 symbols and then apply the MC algorithm. The set of possible modulations remains the same as previous experiments. Three ICA algorithms are tested: fastICA, JADE, and the Gönen algorithm.

It is clear that the fastICA algorithm outperforms the others in terms of percent correct classification. It is interesting to note that both the JADE and Gönen algorithms perform better in the presence of multiple users than when a single user is present. For example when the SNR is 0 dB, the Gönen algorithm performs at 49% correct classification with a single user and 81% correct classification with two additional users. This suggests that both the JADE and Gönen algorithms could be improved for the single user case by artificially adding other signals in the event a single user is detected.

V. Conclusions

The use of BSS techniques in multi-receiver MC applications has not been thoroughly studied. Since BSS techniques introduce a phase ambiguity in the received signals, an algorithm for correcting the phase was introduced and increased MC performance for low SNR. This work shows that the choice of BSS technique requires consideration of both single and multi-user scenarios. For single-user scenarios, the JADE and Gönen algorithms are outperformed by the combination of independent classification decisions. However, the fastICA algorithm performs the best for both single and multi-user scenarios. Used in tandem with an SVM implementation, the fastICA algorithm achieves a percent correct classification of greater than 90 percent for SNR greater than 0 dB. In contrast to optimal likelihood-based algorithms in similar situations, this performance is comparable for significantly less computational complexity.

VI. Acknowledgements

This project was partially supported by the Broadband Wireless Access and Applications Center (BWAC); NSF Award No. 1265960.

References

[1] A. Hyvärinen and E. Oja, “Independent Component Analysis: Algorithms and Applications,” *Neural Networks : the Official Journal of the International Neural Network Society*, vol. 13, no. 4-5, pp. 411–30, 2000.

[2] W. Su, F. Monmouth, J. A. Kosinski, and M. Yu, “Dual-Use of Modulation Recognition Techniques for Digital Communication Signals,” in *Systems, Applications and Technology Conference*, 2006, pp. 1–6.

[3] D. Jakubisin and R. Buehrer, “Improved Modulation Classification Using a Factor-Graph-Based Iterative Receiver,” *MILCOM*, 2012.

[4] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, “A Survey of Automatic Modulation Classification Techniques : Classical Approaches and New Trends,” *Communications, IET*, vol. 1, no. 2, pp. 137–156, 2007.

[5] A. Ramezani-kebrya, I.-m. Kim, D. I. Kim, and R. Inkol, “Likelihood-Based Modulation Classification for Multiple-Antenna Receiver,” vol. 61, no. 9, pp. 3816–3829, 2013.

[6] M. S. Mühlhaus, M. Öner, O. A. Dobre, H. U. Jkel, and F. K. Jondral, “Automatic Modulation Classification for MIMO Systems Using Fourth-Order Cumulants,” *IEEE Vehicular Technology Conference (VTC Fall)*, pp. 1–5, Sep. 2012.

[7] J. L. Xu, W. Su, and M. Zhou, “Distributed Automatic Modulation Classication With Multiple Sensors,” vol. 10, no. 11, pp. 1779–1785, 2010.

[8] J. Cardoso and A. Souloumiac, “Blind Beamforming for non-Gaussian Signals,” *IEE Proceedings-F (Radar and Signal Processing)*, vol. 140, no. 6, 1993.

[9] E. Gonen and J. Mendel, “Applications of Cumulants to Array Processing. III. Blind Beamforming for Coherent Signals,” *IEEE Transactions on Signal Processing*, vol. 45, no. 9, pp. 2252–2264, 1997.

[10] A. Swami and B. Sadler, “Hierarchical Digital Modulation Classification Using Cumulants,” *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 416–429, 2000.

[11] F. Hameed, “On the Likelihood-based Approach to Modulation Classification,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5884–5892, 2009.

[12] A. Hazza and M. Shoaib, “An Overview of Feature-Based Methods for Digital Modulation Classification,” *Conference on Communications, Signal Processing, and their Applications*, vol. 1, no. 08, 2013.

[13] B. E. Boser and et al., “A Training Algorithm for Optimal Margin Classifiers,” in *Proceedings of the 5th annual ACM workshop on Computational Learning Theory*. ACM Press, 1992, pp. 144–152.

[14] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.