A SVM-based Multiuser Detector in CDMA Systems *

Wang Xiangying, Liang Shuangchun and Zhong Yixin

School of Information Engineering
Beijing University of Posts and Telecommunications
Beijing, China 100876
E-mail: ddhyywxy@sina.com.cn

Abstract – As an emerging and very promising learning technique, Support Vector Machine (SVM) has been applied to many classification problems and shows good performance. In this paper, SVM is proposed as a nonlinear multiuser detector (MUD) from a relatively small training data block to detect synchronous CDMA signals in multipath Rayleigh fading channel. We implement SVM as a nonlinear receiver with the radial basis function kernel. The simulation results show that its performance is superior to that of linear MMSE MUD.

I. INTRODUCTION

In wireless CDMA systems, the major performance limits at the receiver are multipath fading and multiple access interference (MAI) caused by co-channel users, which are not orthogonal to the desired user. So the conventional single-user detection is severely affected by the MAI, and a system using conventional detection is interference-limit. Many advanced signal processing techniques have been proposed to combat interference and multipath channel distortion, such as multiuser detection and space-time processing [1][2].

Multiuser detection techniques exploit the underlying structure induced by the spreading waveforms of the DS-CDMA user signals for interference suppression [1]. Verdu developed the optimal maximum-likelihood (ML) detector. The inherent complexity, however, increases exponentially with the number of users, rendering this optimal detector impractical. The linear minimum mean square error (MMSE) multiuser detector (MUD) has been suggested, as it is computationally very simple and can readily be implemented using standard adaptive filter techniques. A more complicated linear detector, however, can only work when the underlying noise-free signal classes are linearly separable. As nonlinear separable cases are common in DS-CDMA channels, support vector machine (SVM) has been considered as a nonlinear MUD [3].

SVM is a new and very promising classification technique developed by Vapnik and his group at AT&T Bell Laboratories. The foundations of SVM are gaining popularity due to many attractive features and promising empirical performance. The formulation embodies the Structural Risk Minimization (SRM) principle, which has been shown to be superior to traditional Empirical Risk Minimization (ERM) principle. SRM minimizes an upper bound on the expected risk, as opposed to ERM that minimizes the error on the training data. It is this difference that equips SVM with a greater ability to generalize, which is the goal in statistical learning [4]. SVM has the advantage over many traditional adaptive learning approaches in terms of performance, complexity and convergence. The main idea behind the technique is to separate the classes with a surface that maximizes the margin between them. In order to find a decision rule with good generalization ability one selects some (small) subset of the training data, called the Support Vectors (SVs). Optimal separation of the SVs is equivalent to optimal separation the entire data. This led to a new method of representing decision functions where the decision functions are a linear expansion on a basis whose elements are nonlinear functions parameterized by the SVs. This type of function representation is especially useful for high dimensional input space: the number of free parameters in this representation is equal to the number of SVs but does not depend on the dimensionality of the space. Another major advantage of using SVM is that a nonlinear SVM can be easily constructed by using kernel functions. Even though we can think of it as a linear algorithm in a high-dimensional space, in practice, it does not involve any computations in that high-dimensional space. By the use of kernels, all necessary computations are performed directly in input space. This is the characteristic twist of SV methods [5].

The SVM technique has been investigated as an adaptive nonlinear MUD [3]. It has been shown that the SVM approach is very effective in multiuser detection system. This paper shows the performance of SVM as a multiuser detector in multipath fading channel with different Doppler frequency `and compares it with linear MMSE MUD.

This paper is organized as follows. Section II describes the DS-CDMA system model. Section III shows the implementation of MMSE MUD. Section IV introduces SVM and proposes a SVM MUD. Some

* This work is supported by the National High Technology Research and Development Program of China (863 Program), No. 2001AA123016

computer simulation results are presented in section V. The paper concludes in section VI.

II. SYSTEM MODEL

We consider a synchronous CDMA system [2] with K users in a frequency selective fading channel, which induces L resolvable multipath components per received signal. Each signal is spread by a random binary spreading sequence of length N. The receiver is assumed equipped with one antenna element. Following a chip matched filter, the discrete-time complex baseband received signal during a given symbol period can be written as a complex N–vector

$$\mathbf{r} = \mathbf{SGAb} + \mathbf{n} . \qquad (1)$$

where $\mathbf{S} = [\mathbf{s}_{1,1} \cdots \mathbf{s}_{1,L}, \cdots, \mathbf{s}_{K,1} \cdots \mathbf{s}_{K,L}]$ is the N-by-KL data spreading code matrix and where $\mathbf{s}_{k,l}$, assumed real, corresponds to the l-th resolvable multipath component of the k-th user's signal. We assume the delay spread is small compared to the symbol period so that inter-symbol interference can be ignored and the delayed code replicas $\mathbf{s}_{k,1} \cdots \mathbf{s}_{k,L}$ for user k can be modeled as L independent random binary spreading codes of length N. The spreading codes are normalized, so that $\| \mathbf{s}_{k,l} \|^2 = 1$, where $\| \cdot \|^2$ denotes the squared Euclidean norm of a vector.

$$\mathbf{G} = diag\left(\begin{bmatrix} g_{1,1} \\ \vdots \\ g_{1,L} \end{bmatrix}, \cdots, \begin{bmatrix} g_{K,1} \\ \vdots \\ g_{K,L} \end{bmatrix} \right)$$

is the defined block diagonal KL-by-K channel matrix, where $g_{K,l}$ is the complex channel coefficient from the k-th user via the l-th multipath component. Matrix $\mathbf{A}$ is the real K-by-K diagonal matrix of amplitudes, $\mathbf{b}$ is the complex K-vector of input data symbols (typically BPSK or QPSK), and $\mathbf{n}$ is the zero-mean complex (circularly symmetric) Gaussian noise vector with independent identically distributed (i.i.d.) components whose real and imaginary components each have variance σ^2.

According to [2], after multipath fading, a sufficient statistic vector for user data is given by K-vector

$$\mathbf{y} = \mathbf{G}^H \mathbf{S}^T \mathbf{SGAb} + \mathbf{G}^H \mathbf{S}^T \mathbf{n} = \mathbf{MAb} + \mathbf{n}_y \qquad (2)$$

where $\mathbf{M} = \mathbf{G}^H \mathbf{S}^T \mathbf{SG}$ is the correlation matrix and $\mathbf{n}_y = \mathbf{G}^H \mathbf{S}^T \mathbf{n}$ is the resulting noise vector. The k-th element of $\mathbf{y}$ is simply the matched filter output for user k, obtained by correlating received signals with its L multipath codes ($\mathbf{s}_{k,1} \cdots \mathbf{s}_{k,L}$), weighting them by the complex conjugate of the corresponding channel ($g_{k,1} \cdots g_{k,L}$) and summing over the multipath indices l.

III. THE MMSE MUD

Given the sufficient statistic vector $\mathbf{y}$ in (2), the MMSE detector applies a linear transformation to $\mathbf{y}$, so that the mean-squared error between the resulting vector and the vector $\mathbf{b}$ is minimized. Hence, the K-by-K MMSE matrix $\mathbf{V}$ should satisfy $\mathbf{V} = \arg \min_{\mathbf{V} \in C^{K \times K}} \{ E \| \mathbf{V}^H \mathbf{y} - \mathbf{b} \|^2 \}$, which results in the following standard Wiener solution [2]:

$$\mathbf{V}^H = \mathbf{A}^{-1} [\mathbf{M} + \sigma^2 \mathbf{A}^{-2}]^{-1} . \qquad (3)$$

We now assume that the adaptive MMSE detector has only knowledge of the K users' codes and multipath timing delays (hence, $\mathbf{S}$ is known) but does not have knowledge of the channel coefficients in $\mathbf{G}$. Then the detector must adaptively derive the linear combiner denoted by $\mathbf{W}$, corresponding to the cascade of the combiners $\mathbf{G}$ and the K-by-K combiner $\mathbf{V}$. The desired matrix $\mathbf{W}$ is a complex-valued KL-by-K matrix which acts on the bank of matched filter outputs and is given by $\mathbf{W}^H = \mathbf{V}^H \mathbf{G}^H$. The MMSE setting $\mathbf{W}$ satisfies $\mathbf{W} = \arg \min_{\mathbf{W} \in K \times KL} \{ E \| \mathbf{W} \mathbf{S}^T \mathbf{r} - \mathbf{b} \|^2 \}$.

In order to compute $\mathbf{W}$, it does not necessarily require explicit channel information. $\mathbf{W}$ can be instead computed iteratively with a multiuser adaptive filtering algorithm [2]. Let the $N_b = 2^K$ possible combinations of

$$\mathbf{b}^{(n)} = [b_1^{(n)}, \cdots, b_K^{(n)}]^T \qquad 1 \le n \le N_b .$$

Define the set of noise-free received signal states as

$$R = \{ \mathbf{x}_n = \mathbf{S}^T \mathbf{SGAb}^{(n)}, \ 1 \le n \le N_b \} .$$

R can be partitioned into two subsets

$$R_{\pm} = \{ \mathbf{x}_n \in R : b_i^{(n)} = \pm 1 \} ,$$

where $b_i^{(n)}$ is the i-th element of $\mathbf{b}^{(n)}$.

If R_+ and R_- are not linearly separable, a linear MUD will have an irreducible error floor even in the noise-free case, as it can only form a hyperplane in the N-dimensional received signal space.

IV. THE SUPPORT VECTOR MACHINE MULTIUSER DETECTOR

A. Introduction of SVM

Consider a binary classification problem where we have two differently labeled finite sets of points in Euclidean n space. Assume that these two sets of points are linearly separable. The goal of SVM (or the optimal margin classifier) is to find the hyperplane that maximizes the minimum distance between any point and the hyperplane [5]. Given a training set $\{ \mathbf{x}_i, y_i \}_{i=1}^N$ with input data $\mathbf{x}_i \in R^n$ and corresponding binary class labels $y_i \in \{-1, +1\}$, the SVM classifier satisfies the following conditions:

$$\begin{cases} \boldsymbol{\omega}^T \mathbf{x}_i + b \ge +1, & if \quad y_i = +1 \\ \boldsymbol{\omega}^T \mathbf{x}_i + b \le -1, & if \quad y_i = -1 \end{cases}$$

which is equivalent to $y_i \left(\boldsymbol{\omega}^T \mathbf{x}_i + b \right) \ge 1, \ i = 1, ..., N$.

Since the margin equals to $\frac{2}{\| \boldsymbol{\omega} \|}$, maximizing the margin is equivalent to minimizing the magnitude of the weights. The problem can then be transformed into a quadratic programming (QP) problem, $\min \frac{1}{2} \| \boldsymbol{\omega} \|^2$ subject to

$$y_i \left(\boldsymbol{\omega}^T \mathbf{x}_i + b \right) \ge 1, \qquad i = 1, ..., N . \qquad (4)$$

SVM can be made more powerful as it can be easily extended

- to construct nonlinear decision functions by introducing kernel function that satisfy Mercers theorem. Here the $\mathbf{x}_i s$ are replaced with the functions $\phi(\mathbf{x}_i)s$ and inner products are replaced with kernel functions $K(\mathbf{x}_i,\mathbf{x}_j)=\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$;
- to handle data that is not separable by introducing slack variables to the QP problem. Data points are penalized if they are misclassified.

The SVM with these modifications can then be described by

$$\min\frac{1}{2}\|\boldsymbol{\omega}\|^2+C\sum_{i=1}^{N}\xi_i$$

subject to

$$y_i(\boldsymbol{\omega}^T\phi(\mathbf{x}_i)+b)\geq 1-\xi_i,\ \ \xi_i\geq 0,\ \ i=1,...,N,\ \ C>0\,.$$

The QP problem can be solved by considering its dual problem where

$$\max Q(\boldsymbol{\alpha})=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j)$$

subject to $0\leq\alpha_i\leq C, i=1,...,N$ and $\sum_{i=1}^{N}\alpha_i y_i=0$

where $\boldsymbol{\alpha}=(\alpha_1\cdots\alpha_N)^T$. The decision function can then be written as

$$f(\mathbf{x})=sign\left(\sum_{i=1}^{N}\alpha_i y_i K\left(\mathbf{x},\mathbf{x}_i\right)+b\right).$$

B. SVM Multiuser Detector

In general, the receiver can have access to a block of M training samples $\{\mathbf{x}(j),b_i(j)\}_{j=1}^{M}$ for user i. For notational convenience, denote the training set of M noisy received signal vectors as

$$X=\{\mathbf{x}_j=\mathbf{x}(j),\ \ 1\leq j\leq M\}$$

and the set of corresponding class labels as

$$Z=\{z_j=b_i(j),\ \ 1\leq j\leq M\}\,.$$

Applying the standard SVM method, a SVM detector can be constructed for user i

$$f_{SVM}(\mathbf{x}_m)=\sum_{j=1}^{M}\alpha_j z_j\cdot K(\mathbf{x}_m,\mathbf{x}_j)+b$$

where the set of Lagrangian multipliers $\{\alpha_j\}$, is the solution of the QP problem

$$\boldsymbol{\alpha}=\arg\min_{\boldsymbol{\alpha}}\left\{\frac{1}{2}\sum_{j=1}^{M}\sum_{l=1}^{M}\alpha_j\alpha_l z_j z_l\cdot K(\mathbf{x}_j,\mathbf{x}_l)-\sum_{j=1}^{M}\alpha_j\right\}$$

with the constraints

$$0\leq\alpha\leq C,\ \ 1\leq j\leq M\ \text{ and }\ \sum_{j=1}^{M}\alpha_j z_j=0\,.$$

In this particular application, Chang Chih-Chung's LIBSVM toolkit is used in Matlab6.5 [6]. And it was found advantageous to choose the radial basis function (RBF) kernel of: $K(\mathbf{x}_j,\mathbf{x}_l)=\exp\left(-\gamma\left\|\mathbf{x}_j-\mathbf{x}_l\right\|^2\right)$. The offset constant b is usually determined from the so-called "margin" SVs. Because the optimal decision boundary passes through the origin of the received signal space and possesses certain symmetric properties due to the symmetric structure of R_+ and R_-, b=0 can be used. The user-defined parameter C controls the tradeoff between model complexity and training error. In our application, we will choose C empirically.

The set of SVs, denoted by X_{SVM}, is given by those $\mathbf{x}_j$s with nonzero Lagrangian multipliers $0<\alpha_j<C$. X_{SVM} is usually a small subset of the training data set X. Thus the SVM MUD requires computing the decision variable

$$f_{SVM}(\mathbf{x}_m)=\sum_{\mathbf{x}_j\in X_{SVM}}\alpha_j z_j\exp\left(-\gamma\left\|\mathbf{x}_m-\mathbf{x}_j\right\|^2\right)$$

and making the decision of user i data with

$$\hat{b}_i(m)=sign(f_{SVM}(\mathbf{x}_m))\,.$$

V. SIMULATION RESULTS

In this section, we will investigate the performance of SVM multiuser detector in multipath Rayleigh fading channel. We choose the "Channel B" outdoor to indoor and pedestrian test environment tapped-delay-line parameters channel defined by ITU-R M.1225 for the evaluation of radio transmission technologies for IMT-2000 [7]. The multipath time delays and the variance of this channel are shown in Table I. The maximum Doppler frequency is set to be 10Hz.

In this simulation, after spreading spectrum the chip rate is 3.84Mc/s, spread factor is 63, and the multipath number L for each user is 6. MMSE MUD method is computed with the iteratively adaptive filtering algorithm. To construct a SVM MUD, 600 training data are generated for each SNR and K. After SVM is trained, 10000 user data are received. We choose the different parameter γ and C to compare the influence on system performance.

We can see from table II that SVM kernel function parameter γ has some influence on the construction of SVM model. If γ is too large or small, the bit error rate (BER) and the number of SV will increase.

It can be seen from table III that, when C decreases, the number of SV increases, and the system BER performance is improved. The reason is that when choosing less SVs, SVM actually constructs the separating hyperplane only based on the outliers, which unfortunately are the training samples contaminated with the largest noise, so it is helpful to use more SVs. However, if C is too small, SVM overfits by taking too many SVs which results in longer training time and a higher error rate, we need to take these factors into consideration when choosing C.

Fig. 1 and Fig. 2 show the performance of MMSE and SVM MUD when Doppler frequency is 10Hz, C is 0.1, and $\gamma=0.075$. We can see that the performance of SVM MUD is superior to that of MMSE MUD at different user number and SNR.

VI. CONCLUSION

This paper discusses how to implement a SVM-based CDMA multiuser detector. This method requires relatively small training sequences. Simulation results show that the performance of SVM MUD is superior to that of MMSE MUD. The further study is how to incorporate the sample-by-sample adaptive training methodology into the SVM approach. Because the optimization technique associated with the SVM method requires a considerable amount of computation, especially for a large *M*, efficient SVM implementation is also our goal in further research.

TABLE I. Characteristics of the ITU-R M.1225 "Channel B"

Relative delay (ns)	Average power (dB)
0	0
200	-0.9
800	-4.9
1200	-8.0
2300	-7.8
3700	-23.9

TABLE Ⅱ. Influence of γ on constructing the SVM MUD for SNR = 0dB, K = 20, C = 0.1, Doppler frequency = 10Hz

γ	Avg. of BER	No. of SV
1	0.13241	167 to 216
0.75	0.13061	145 to 195
0.5	0.1292	127 to 173
0.125	0.12885	111 to 158
0.075	0.12939	123 to 175
0.01	0.13404	216 to 274

TABLE Ⅲ. Influence of C on constructing the SVM MUD for SNR=0 dB, K=20, $\gamma = 0.075$, Doppler frequency = 10Hz

C	Avg. of BER	No. of SV
5	0.13066	52 to 106
2	0.12904	58 to 107
1	0.12792	64 to 112
0.75	0.12781	67 to 114
0.1	0.12885	111 to 158
0.05	0.12974	142 to 194

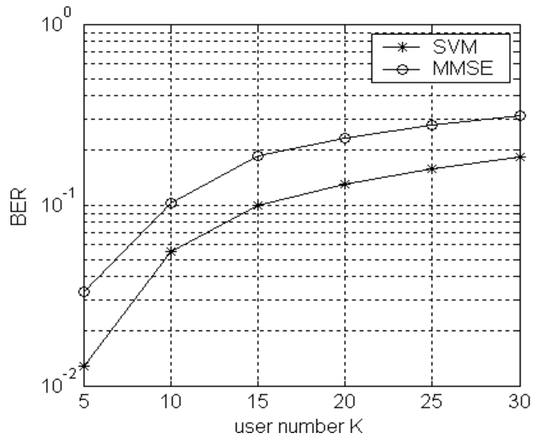

Fig. 1. Performance of SVM MUD and MMSE MUD with different user number, SNR = 0 dB, Doppler frequency =10Hz, C = 0.1, $\gamma = 0.075$

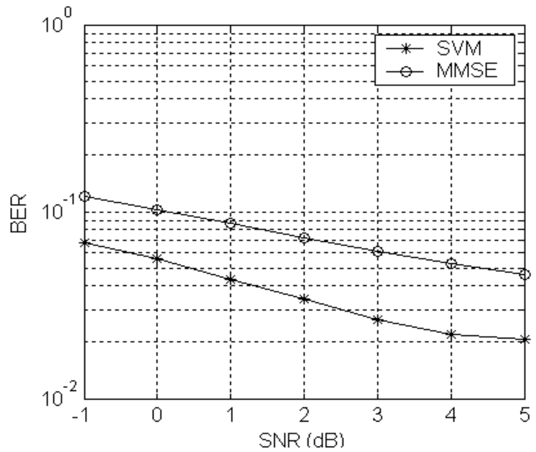

Fig. 2. Performance of SVM MUD and MMSE MUD with different SNR, K = 10, Doppler frequency = 10Hz, C = 0.1, $\gamma = 0.075$

REFERENCES

[1] S. Verdu, *Multiuser Detection*, Cambridge, U.K.: Cambridge University Press, 1998.

[2] C.B. Papadias and H. Huang, "Linear space-time multiuser detection for multipath cdma channels", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 2, pp. 254-264, Feb. 2001.

[3] S. Chen, A.K. Samingan and L. Hanzo, "Support vector machine multiuser receiver for DS-CDMA signals in multipath channels", *IEEE Trans. on Neural Networks*, vol.12, no.3, pp. 604-611, May 2001.

[4] S.R. Gunn, "Support Vector Machines for Classification and Regression", *ISIS Technical Report*, University of Southampton, May 1998.

[5] V. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.

[6] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: A Library for Support Vector Machines, 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[7] ITU-R Rec. M.1225, "Guideline for evaluation of radio transmission technologies for IMT-2000".