

A hybrid generative/discriminative method for EEG evoked potential detection

Yonghong Huang, Misha Pavel

Kenneth E. Hild

Oregon Health & Science University

Department of Science & Engineering

huang@csee.ogi.edu, pavel@bme.ogi.edu, k.hild@ieee.org

Deniz Erdogmus

Northeastern University

Department of ECE

409 Dana Research Center

erdogmus@ece.neu.edu

Santosh Mathan

Honeywell Laboratories

Human Centered Systems Group

Redmond, WA 98052

Santosh.Mathan@honeywell.com

Abstract—We propose a new method for the detection of evoked potentials that combines a generative model and a discriminative classifier. The method is a variant of the support vector machine (SVM), which uses the Fisher kernel. The kernel function is derived from a generative statistical model known as mixed effects model (MEM). Instead of arbitrarily selecting the Gaussian kernel for the SVM, we exploit the Fisher kernel derived from the MEM for the SVM. The strength of this approach is that it combines the rich information encoded in the generative model, the MEM, with the discriminative power of the SVM algorithm. Our results show that the new method of combining the two complementary approaches – the generative model (MEM) and the discriminative model (SVM) via the Fisher kernel – achieves substantial improvement over the generative model (MEM) and provides better performance than the discriminative model (Gaussian kernel SVM) on the detection of evoked potentials in neural signals.

I. INTRODUCTION

Generative and discriminative learning approaches are two prevailing and powerful, yet different, paradigms in machine learning. Generative learning models, such as Bayesian inference [1] attempt to model the underlying distributions of the variables in order to compute classification and regression functions. These methods provide a rich framework for learning from prior knowledge. Discriminative learning models, such as support vector machines (SVM) [2] avoid generative modeling by directly optimizing a mapping from the inputs to the desired outputs by adjusting the resulting classification boundary. These latter methods commonly demonstrate superior performance in classification. Recently, researchers have investigated the relationship between these two learning paradigms and have attempted to combine their complementary strengths [3]. In this work we combine the Fisher kernel technique, which is a generative model, with a discriminative classifier. Jaakkola et al. first introduced the Fisher kernel to create a generic mechanism for incorporating generative probability models, such as hidden Markov models, into discriminative models, such as SVMs. This was done in an attempt to overcome the problems associated with variable-length feature vectors [4]. Jaakkola and his colleagues showed that combining the Fisher kernel with a discriminative model significantly improved performance over a hidden Markov model for detecting remote protein homologies [5]. Mika et al. developed a fast training algorithm that allows the Fisher

kernel algorithm to be applied to large datasets [6]. Tsuda et al. analyzed the statistical properties of the Fisher kernel [7]. The Fisher kernel has been widely used in many areas, such as large scale web audio classification [8] and speech recognition [9].

To improve the performance of single-trial electroencephalography (EEG) event related potential (ERP) detection [10], we propose a new ERP detector that combines a generative model and a discriminative model. More specifically, our method uses the Fisher kernel derived from the mixed effects model (MEM) in the SVM. A MEM [11] is a statistical hierarchical model. It was first proposed for the analysis of longitudinal time-series data [12] and has been widely used for the analysis of longitudinal data and image shape analysis [11], [13]. There are two sources of variation in the MEM: between-individual variations and within-individual variation. An advantage of the MEM is the ability to combine the data by introducing multilevel random effects. Therefore, it is well suited for the analysis of longitudinal data and biomedical data. In our previous study [14], we applied this statistical approach to the classification of single-trial multichannel EEG sequences. We showed that the MEM models the neural responses to target stimuli and distractor stimuli quite well and that the performance of the MEM approaches that of the Gaussian kernel SVM (GKSVM) for ERP detection. Since it incorporates multilevel random effects, the MEM easily handles data with multiple sources of variation, such as EEG data. For example, we can easily use the population-averaged parameters to specify the EEG signal that is common across subjects and/or sessions and we can use the subject-specific parameters to specify the signal that is unique from one subject/session to the next. This allows us to easily aggregate data across multiple subjects/sessions. Consequently, the MEM provides a principled mechanism for combining historical and new data, which greatly simplifies online adaptation.

In this work, we demonstrate the feasibility of the proposed Fisher kernel SVM (FKSVM) for single-trial ERP detection. Instead of arbitrarily selecting the Gaussian kernel for the SVM, we want to use the Fisher kernel derived from the MEM. The strength of this approach is that it combines the rich biological information encoded in the MEM with the discriminative power of the SVM algorithm. Our results show

that the new method of combining the two complementary approaches – the generative model (MEM) and the discriminative model (SVM) via the Fisher kernel – achieves substantial improvement over using either the generative model (MEM) or the discriminative model (GKSVM) alone for single-trial ERP detection.

II. MODELS

A. Fisher Kernel

The Fisher kernel provides a natural similarity measure between data samples, which takes into account the underlying probability distribution. The Fisher kernel operates in the gradient space of the generative model, specifically the gradient of the loglikelihood with respect to the parameters. It describes how the parameters contribute to the process of generating a particular example. For any data vector \mathbf{y}_i and model parameters $\boldsymbol{\theta}$, the Fisher score is a row vector and which is defined as

$$\mathbf{U}_{\mathbf{y}_i} = \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_i | \boldsymbol{\theta}). \quad (1)$$

The Fisher Information matrix is defined as

$$\mathbb{I} = E_{\mathbf{y}_i} \{ \mathbf{U}_{\mathbf{y}_i}^T \mathbf{U}_{\mathbf{y}_i} \}, \quad (2)$$

where $E_{\mathbf{y}_i} \{ \}$ is the expectation over $p(\mathbf{y}_i | \boldsymbol{\theta})$. The Fisher kernel is defined as

$$K(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{U}_{\mathbf{y}_i} \mathbb{I}^{-1} \mathbf{U}_{\mathbf{y}_j}^T, \quad (3)$$

where \mathbf{y}_i and \mathbf{y}_j are two data samples. The Fisher kernel compares two data samples through the directions in which they stretch the model parameters. If the gradient vectors are similar, then the two samples would adapt the model in the same way. Detailed information and properties of the Fisher kernel can be found in Jaakkola's paper [4] and Tsuda's paper [7].

B. Mixed Effects Models

The general MEM model for individual i of N , $i = 1, \dots, N$ is written as,

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i. \quad (4)$$

where

- \mathbf{y}_i is the $n \times 1$ vector of observations for the i th individual and n is the number of observations for the i th individual,
- $\boldsymbol{\alpha}$ is the $p \times 1$ population fixed effects vector, and \mathbf{X}_i is the $n \times p$ population design matrix of fixed effects,
- \mathbf{b}_i is the $k \times 1$ individual random effects vector; the random effects are assumed to be independent and have a normal distribution $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$, where \mathbf{D} is a $k \times k$ positive definite covariance matrix; and \mathbf{Z}_i is the $n \times k$ individual design matrix of random effects,
- $\boldsymbol{\varepsilon}_i$ is an $n \times 1$ vector of independent and identically distributed (iid) errors with zero mean and positive definite covariance matrix, which represents the within-individual variance, and the error terms are assumed to have a normal distribution $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, where \mathbf{I}_n denotes an $n \times n$ identity matrix.

This model (4) can be written as a multivariate normal distribution, $\mathbf{y}_i \sim N(\mathbf{X}_i \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T)$. Detailed information about the parameter estimation and design matrix construction can be found in [14]. We let $Var(\mathbf{y}_i)$ represent $\mathbf{V}_i = \sigma^2 \mathbf{I}_n + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T$, if the covariance parameters $\hat{\sigma}^2$ and $\hat{\mathbf{D}}$, were known, then \mathbf{V}_i is also known and we could estimate $\boldsymbol{\alpha}$ and \mathbf{b}_i . Assuming \mathbf{y}_i is independent for each i , the joint density function of \mathbf{y}_i is

$$p(\mathbf{Y}; \boldsymbol{\theta}) = \prod_{i=1}^N \frac{\exp[-\frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})]}{(2\pi)^{\frac{n}{2}} |\mathbf{V}_i|^{\frac{1}{2}}}, \quad (5)$$

where $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ and $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \mathbf{D}, \sigma)$. The log-likelihood function for the MEM is given by

$$\begin{aligned} l(\boldsymbol{\theta}) &= -\frac{1}{2} \left\{ \left(\sum_{i=1}^N n \right) \ln(2\pi) \right. \\ &\quad \left. + \sum_{i=1}^N [\ln |\mathbf{V}_i| + (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha})] \right\} \end{aligned} \quad (6)$$

III. FISHER KERNEL DERIVED FROM THE MEM

The key idea here is to derive the Fisher kernel function from the MEM and use it to replace the Gaussian kernel in the SVM for ERP detection. First we train a MEM and use the MEM to map each new example into its Fisher score. Then we compute the Fisher kernel function on the basis of the Euclidean distance between the scores of the new sample and the training samples; Finally we can measure the discriminant value for the new example from the SVM discriminant function.

A. Fisher scores derived from the MEM

We define the MEM Fisher score as

$$\begin{aligned} \mathbf{U}_{\mathbf{y}_i} &= \nabla_{\boldsymbol{\theta}} \log p(\mathbf{y}_i | \boldsymbol{\theta}) \\ &= [\nabla_{\boldsymbol{\alpha}} \log p(\mathbf{y}_i | \boldsymbol{\alpha}), \nabla_{\mathbf{D}} \log p(\mathbf{y}_i | \mathbf{D}), \nabla_{\sigma} \log p(\mathbf{y}_i | \sigma)], \end{aligned} \quad (7)$$

where the model parameters of the MEM $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \mathbf{D}, \sigma)$ and data samples $\mathbf{y}_i \sim N(\mathbf{X}_i \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n + \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^T)$, for $i = 1, 2, \dots, N$ as in Equation (1) and (2). We use an MEM trained from target examples to model a given ERP and we derive the Fisher score based on the target MEM. This intuition is that we would like to enhance the discriminative power of the classifier.

1) *Fisher scores of parameter $\boldsymbol{\alpha}$* : Fisher scores respective to the fixed effect parameter of the MEM $\boldsymbol{\alpha}$ is a $1 \times p$ row vector

$$\frac{\partial l}{\partial \boldsymbol{\alpha}^T} = \left[\frac{\partial l}{\partial \alpha_1}, \dots, \frac{\partial l}{\partial \alpha_m}, \dots, \frac{\partial l}{\partial \alpha_p} \right] \quad (8)$$

where $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_m, \dots, \alpha_p]^T$ is a column vector. Based on (6) and if we let $\mathbf{a} = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\alpha}$, we have

$$\begin{aligned} \frac{\partial l}{\partial \alpha_m} &= -\frac{1}{2} \frac{\partial (\mathbf{a}^T \mathbf{V}_i^{-1} \mathbf{a})}{\partial \alpha_m} \\ &= \mathbf{a}^T \mathbf{V}_i^{-1} \mathbf{X}_{i,m} \end{aligned} \quad (9)$$

Since \mathbf{V}^{-1} is a symmetric matrix, we have

$$\frac{\partial l}{\partial \boldsymbol{\alpha}^T} = (\mathbf{y}_i - \mathbf{X}\boldsymbol{\alpha})^T \mathbf{V}^{-1} \mathbf{X} \quad (10)$$

2) *Fisher scores of parameter D*: The covariance matrix D of the random effects of the MEM is a $k \times k$ matrix, Based on (6), the Fisher scores respective to parameter D is a $k \times k$ matrix,

$$\frac{\partial l}{\partial \mathbf{D}} = -\frac{1}{2} \left[\frac{\partial \ln |\mathbf{V}|}{\partial \mathbf{D}} + \frac{\partial (\mathbf{a}^T \mathbf{V}^{-1} \mathbf{a})}{\partial \mathbf{D}} \right]. \quad (11)$$

For each entry (m,l) of matrix D , we have

$$\frac{\partial l}{\partial \mathbf{D}_{ml}} = -\frac{1}{2} \left[\frac{\partial \ln |\mathbf{V}|}{\partial \mathbf{D}_{ml}} + \frac{\partial (\mathbf{a}^T \mathbf{V}^{-1} \mathbf{a})}{\partial \mathbf{D}_{ml}} \right]. \quad (12)$$

The first part of (12) is

$$\begin{aligned} \frac{\partial \ln |\mathbf{V}|}{\partial \mathbf{D}_{ml}} &= \sum_{ij} \frac{\partial \ln |\mathbf{V}|}{\partial \mathbf{V}_{ij}} \cdot \frac{\partial \mathbf{V}_{ij}}{\partial \mathbf{D}_{ml}} \\ &= \sum_{ij} (\mathbf{V}^{-1})_{ij} \cdot (\mathbf{Z} \cdot \mathbf{E}_{ml} \cdot \mathbf{Z}^T)_{ij}, \end{aligned} \quad (13)$$

Where \mathbf{E}_{ij} is elementary matrix with only nonzero entry of 1 occurring at location (i,j). The second part of (12) is

$$\frac{\partial (\mathbf{a}^T \mathbf{V}^{-1} \mathbf{a})}{\partial \mathbf{D}_{ml}} = -\mathbf{a}^T (\mathbf{V}^{-1} \cdot \mathbf{Z} \cdot \mathbf{E}_{ml} \cdot \mathbf{Z}^T \cdot \mathbf{V}^{-1}) \mathbf{a} \quad (14)$$

Base on (13) and (14), (12) can be written as

$$\begin{aligned} \frac{\partial l}{\partial \mathbf{D}_{ml}} &= -\frac{1}{2} \left[\sum_{ij} (\mathbf{V}^{-1})_{ij} \cdot (\mathbf{Z} \cdot \mathbf{E}_{ml} \cdot \mathbf{Z}^T)_{ij} \right. \\ &\quad \left. - \mathbf{a}^T (\mathbf{V}^{-1} \cdot \mathbf{Z} \cdot \mathbf{E}_{ml} \cdot \mathbf{Z}^T \cdot \mathbf{V}^{-1}) \mathbf{a} \right]. \end{aligned} \quad (15)$$

Therefore (11) can be written as

$$\frac{\partial l}{\partial \mathbf{D}} = \sum_{ij} (\mathbf{E}_{ml} \cdot \frac{\partial l}{\partial \mathbf{D}_{ml}}) \quad (16)$$

3) *Fisher scores of parameter σ^2* : Because the variance of the noise of the MEM, σ^2 , is a scalar, the Fisher scores for the parameter σ^2 is a scalar. Based on (6), we have

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2} \left[\frac{\partial \ln |\mathbf{V}|}{\partial \sigma^2} + \frac{\partial (\mathbf{a}^T \mathbf{V}^{-1} \mathbf{a})}{\partial \sigma^2} \right]. \quad (17)$$

The first part of (17) is

$$\frac{\partial \ln |\mathbf{V}|}{\partial \sigma^2} = \text{tr}(\mathbf{V}^{-1}). \quad (18)$$

The second part of (17) is

$$\frac{\partial (\mathbf{a}^T \mathbf{V}^{-1} \mathbf{a})}{\partial \sigma^2} = -\mathbf{a}^T \cdot (\mathbf{V}^{-1})^2 \cdot \mathbf{a} \quad (19)$$

Therefore we can write (17) as

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2} [\text{tr}(\mathbf{V}^{-1}) - \mathbf{a}^T \cdot (\mathbf{V}^{-1})^2 \cdot \mathbf{a}] \quad (20)$$

We can now calculate the Fisher score with respective to each parameter using (10),(16) and (20). Based on (7), we can obtain the complete Fisher score for each data sample as

$$\mathbf{U}_{\mathbf{y}_i} = \left[\frac{\partial l}{\partial \boldsymbol{\alpha}^T}, \text{Vec} \left(\frac{\partial l}{\partial \mathbf{D}} \right), \frac{\partial l}{\partial \sigma^2} \right], \quad (21)$$

where $\text{Vec}(\cdot)$ is the vectorization of a matrix.

B. Fisher kernel derived from the MEM

After we have the Fisher scores, we can construct the Fisher kernel based on (3). We use a natural approximation, which involves sample averaging for the Fisher information matrix on training and test respectively,

$$\hat{\mathbb{I}}_{tr} = \frac{1}{n} \sum_{k=1}^n \mathbf{U}_{\mathbf{y}_k^{tr}} \mathbf{U}_{\mathbf{y}_k^{tr}}^T. \quad (22)$$

Similarly we obtain $\hat{\mathbb{I}}_{te}$. We compute the Fisher kernel from the training data for the SVM training using,

$$K(\mathbf{y}_i^{tr}, \mathbf{y}_j^{tr}) = \frac{1}{s} \mathbf{U}_{\mathbf{y}_i^{tr}} \hat{\mathbb{I}}_{tr}^{-1} \mathbf{U}_{\mathbf{y}_j^{tr}}^T, \quad (23)$$

where \mathbf{y}_i^{tr} and \mathbf{y}_j^{tr} are two training data samples and s is scaling constant. We compute the Fisher kernel from the test data and the support vectors for the test,

$$K(\mathbf{y}_i^{te}, \mathbf{y}_j^{sv}) = \frac{1}{s} \mathbf{U}_{\mathbf{y}_i^{te}} \hat{\mathbb{I}}_{te}^{-1} \mathbf{U}_{\mathbf{y}_j^{sv}}^T, \quad (24)$$

where \mathbf{y}_i^{te} is one test data sample and \mathbf{y}_j^{sv} is one support vector. After we have the Fisher kernels, we can use them to replace the Gaussian kernels in the SVM.

IV. RESULTS

We evaluated the proposed approach, which we refer to as FKSVM, on single-trial ERP detection. Six professional image analysts participated in the experiment. The subjects performed target detection and their EEG signals were recorded. We eliminated one subject because he finished only one test session. Four subjects had seven test sessions and one subject had four test sessions. The 32 channel EEG data were collected and preprocessed as in [15]. The performance is estimated using the area under the receiver operating characteristic (ROC) curve (AUC) [16]. We use Delongs non-parametric approach [17] to compare two correlated AUCs by generating an estimated covariance matrix. Our null Hypothesis H_0 is that there is not statistically significant difference between the two correlated AUCs. We used 10-fold cross validation [1] to select the optimal parameters for each subject (the scaling constant s and cost parameter C of the FKSVM, the kernel size σ and cost parameter C for the GKSVM, and the parameters of the design matrices for the MEM).

A. Comparison of the FKSVM with the MEM

The non-parametric test on two correlated AUC measurements from the FKSVM and the MEM classifiers shows that the FKSVM achieves significantly better performance than the MEM. The ROC curves of the FKSVM and the MEM for the 32 test sessions of the five subjects are shown in Figure 1. This figure shows that the FKSVM achieves substantial improvement over the MEM in single-trial ERP detection. The averaged AUCs across five subjects for the FKSVM and the MEM are 0.89 and 0.85, respectively. The two-tailed p values for five subjects $\ll 0.05$ ($p = 0.008, 0.004, 0.001, 0.018$ and 0.025 respectively). One can conclude that the difference between the AUCs from the FKSVM and the MEM is statistically significant and reject H_0 .

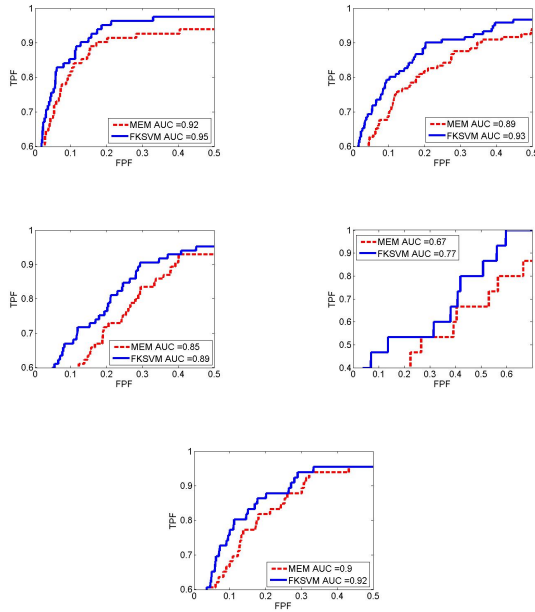


Fig. 1. The test ROC curves of the FKSVM and the MEM for five subjects (congregate multiple test session outputs for each subject)

TABLE I

TEST PERFORMANCE OF THE FKSVM AND GKSVM FOR FIVE SUBJECTS

Subject	FKSVM	GKSVM
1	0.95	0.94
2	0.93	0.92
3	0.89	0.85
4	0.77	0.72
5	0.92	0.92
Mean	0.89	0.87

B. Comparison of the FKSVM with the GKSVM

The AUCs of the FKSVM and the GKSVM of the five subjects are shown in Table I. This table shows that the FKSVM performs better than the GKSVM on single-trial ERP detection. The AUCs across five subjects for the FKSVM and the GKSVM are 0.89 and 0.87, respectively. The two-tailed p values for five subjects > 0.05 . Therefore the difference between the AUCs from the FKSVM and the GKSVM are not statistically significant at 95% confidence level.

V. DISCUSSION

We begin with a trained MEM and use this MEM to map each new data vector we wish to detect into its Fisher score. We then compute the Fisher kernel function on the basis of the Euclidean distance between the score vector for the new ERP example and the score vectors for training examples. Next we measure the discriminant value for this new ERP example from the discriminant function of the SVM. Jaakkola suggests [4] using the Fisher kernel found by setting the Fisher information matrix to an identity matrix for computational simplicity. However our data analysis show that the sample mean approximation in (22) for the Fisher information matrix

produces better performance than an identity matrix. The performance of the FKSVM was compared with that produced by the generative model (MEM) and a competing discriminative model (GKSVM). The results show that the proposed method, which combines the MEM and SVM via the Fisher kernel, outperforms both the MEM and the SVM used in isolation for single-trial ERP detection in our data. This general approach of combining a generative model with a discriminative method may have applications in other areas of biosignal classification and analysis as well.

ACKNOWLEDGMENT

The authors would like to thank Zhengdong Lu for the discussion. This work was supported by DARPA and by NSF under grants *ECS* – 0524835, *ECS* – 0622239, and *IIS* – 0713690. Satellite images included in this paper were produced by DigitalGlobe Inc., Longmont, CO 80501, USA. (c) 2003

REFERENCES

- [1] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, Wiley, Inc., New Jersey, NY, 2001.
- [2] V.N. Vapnik, *Statistical learning theory*, John Wiley & Sons, Inc., New York, NY, 1998.
- [3] T. Jebara, *Machine Learning Discriminative and Generative*, Kluwer Academic Publishers Group, Dordrecht, Netherlands, 2004.
- [4] T. Jaakkola and D. Haussler, “Exploiting generative models in discriminative classifiers,” in *Proc. of the Conf. on Advances in neural information processing systems*, Denver, CO, 1998, pp. 487–493.
- [5] T. Jaakkola, M. Diekhans, and D. Haussler, “A discriminative framework for detecting remote protein homologies,” *Journal of Computational Biology*, vol. 7, no. 1-2, pp. 95–114, 2000.
- [6] S. Mika, A. Smola, and B. Scholkopf, “An improved training algorithm for kernel fisher discriminants,” in *Proc. of the 8th Int. Workshop on Artificial Intelligence and Statistics*, Key West, Florida, 2001, pp. 98–104.
- [7] K. Tsuda, S. Akaho, M. Kawanabe, and K.R. Muller, “Asymptotic properties of the fisher kernel,” *Journal of Neural Computation*, vol. 16, no. 1, pp. 115137, 2004.
- [8] P.J. Moreno and R. Rifkin, “Using the fisher kernel method for web audio classification,” in *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, 2000, pp. 2417–2420.
- [9] N. Smith and M. Gales, “Speech recognition using svms,” in *Proc. of the Conf. on Advances in neural information processing systems*, Vancouver, Canada, 2002, pp. 1197–1204.
- [10] S.J. Luck, *An Introduction to the Event-Related Potential Technique*, The MIT Press, Cambridge, MA, 2005.
- [11] E. Demidenko, *Mixed Models Theory and Application*, John Wiley & Sons, Inc., New Jersey, NY, 2004.
- [12] N.M. Laird and J.H. Ware, “Random-effect models for longitudinal data,” *Biometrics*, vol. 38, pp. 963–974, 1982.
- [13] Z. Lu, *Constrained clustering and cognitive decline detection*, PhD thesis, Oregon Health & Science University, Portland, OR, 2008.
- [14] Y. Huang, D. Erdogmus, S. Mathan, and M. Pavel, “Mixed effects models for eeg evoked response detection,” in *Proc. of IEEE Workshop on Machine Learning for Signal Processing*, Cancun, Mexico, 2008.
- [15] Y. Huang, D. Erdogmus, S. Mathan, and M. Pavel, “Large-scale image database triage via eeg evoked responses,” in *Proc. of IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Las Vegas, NV, 2008, pp. 429–432.
- [16] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, New York, NY, 2003.
- [17] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, “Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach,” *Biometrics*, vol. 44, pp. 837–845, 1988.