AMELIORATION OF PHYSICAL ACTIVITY ESTIMATION FROM ACCELEROMETER SENSORS USING PRIOR KNOWLEDGE

A. Ataya, P. Jallon

CEA-Leti Minatec Campus
17 rue des Martyrs, 38054 Grenoble, France

ABSTRACT

Human physical activity assessment using inertial sensor's data has become a prominent research area in the biomedical engineering field and an important application area for pattern recognition. This paper proposes to improve physical activity detection by combining prior knowledge concerning activity sequences with predictions of a support vector machine classifier (SVM). The temporal stable nature of activities is modeled by a directed graph Markov chain to reinforce decisions obtained using activity classes' confidence measures of a traditional SVM. We therefore review existing approaches dealing with determining these confidence measures for SVM classification. We then propose new methods for confidence measures estimation for SVM bi-class and multi-class problems. While applying the graph with proposed techniques for confidence estimation, results show superlative recognition rate of 92% for classifying 6 activities from data collected by a tri-axial accelerometer worn on belt.

Index Terms— Physical activity, inertial sensors, accelerometers, pattern recognition, Markov chain, SVM, prior information, confidence measures.

1. INTRODUCTION

Human physical activity recognition has been receiving considerable interest in recent years due to its potential application in an extensive range of fields including biomechanics, tracking of physically and mentally disabled persons, monitoring of elderly and medical diagnosis especially for quantifying the relationship between physical activity behavior and chronic diseases as cardiovascular disease, diabetes, and certain cancer types [1] and also its important role in reducing the risk of obesity [2]. Despite the fact that numerous studies have applied computer vision for this purpose, the use of wearable miniaturized inertial sensors such as accelerometers has proved certain favorable advantages [3].

The great majority of studies attempting to categorize physical activities using body-worn sensors' data employ pattern recognition for the classification [3, 4, 5, 6]. In a general study, [7] justified the advantage of using discriminative classification methods over generative ones. Among

the discriminative methods, large margin classifiers have become prominent due to the eminent performance of support vector machines (SVM) in many real world problems dealing with classification. Nevertheless, most classifier learning techniques including SVM assume implicitly that the continuous data collected from inertial sensors are obtained from time-independent practiced sequence of activities. In real life, this assumption is not likely to be always valid and different practiced activities in the same time range are sequentially correlated. For instance, human physical activities have a nonfluctuating nature in time and a performed activity at a specific instant is more likely to be the same as the one practiced at the previous instant. This type of prior knowledge on the sequence of activities serves as a crucial complement to the classifier's decision for improving the classified activities results. In this case, the classifier constitutes just a part of the overall decision and confidence measurements can be of great usefulness for introducing this prior knowledge. In this context, a new decision technique established from Markov modeling approach was developed under the name of graph based methods in [8, 9] for several classifiers. In this study, we propose an efficient combination of graph based methods with a new method of evaluation of confidence measures from SVM outputs based on estimating posterior probabilities capable of defining the degrees of membership of observed data to different activity classes.

In section 2 graph based methods are presented to formalize the problem of merging physical activity-related prior knowledge with the classifiers' judgment in an overall decision technique. In section 3 the base SVM classifier is introduced as well as the different existing methods for the estimation of confidence measures in the belonging of data points to different classes. Novel methods of confidence estimation for the binary classification as for the multi-class case are discussed and justified. Data collection for evaluation of the proposed methods are presented in 4. In section 5 performance results are discussed and conclusions are drawn.

2. GRAPH BASED METHODS

Human daily activities exhibit generally significant sequential correlation ; in the sequence of performed activities,

nearby activity classes are likely to be time-related to each other. Despite this fact, most classifier-learning algorithms assume that the data collected by inertial sensors are drawn from a time-independent sequence of activities. Consequently, if $\mathcal{O}_m$ is the observation vector containing the features used for activity estimation at instant m, the conventional supervised classification methods make a prediction $\hat{A}_m$ of the real practiced activity A_m while basing their decisions exclusively on the observation $\mathcal{O}_m$. Graph based methods introduced in [8, 9] remedy this incomplete time-independence assumption and reestablish a more realistic one. The principal idea is to introduce prior knowledge concerning the searched sequence of activities by considering that this latter is correlated in time rather than being time-independent. Using graph based methods, the decision concerning the practiced activity at time index m is made by taking into account both the observation vector $\mathcal{O}_m$ and the classification history of observations corresponding to previous time indexes. Formally speaking, if N is the total number of activities and $\{A_m\}_m$ represents the sequence of activities, temporal dependencies can be modeled using a directed graph described by a Markov chain, and are therefore entirely characterized by the initialization probabilities $p(A_1 = i)$ $(i = 1, 2, \ldots, N)$, and the transition probability distribution $p(A_m = j | A_{m-1} = i)$.

The infeasibility of a direct shift from an activity i to activity j can be modeled by assigning the corresponding transition a null probability. On the other hand, a person performing certain activity at a specific time instant has a tendency to carry out the same activity at the following instant. In the graph model, this type of prior knowledge can be comprised by attributing self transition probabilities values sufficiently close to unity to favor the temporal stable nature of the different executed activities. In the absence of any prior assumption about the initial practiced activity by the person, the different activities at instant 1 are considered to be equiprobable. The author in [9] demonstrated that applying the graph to a classifier based on Bayesian modeling of data amounts to modeling the couple of sequences $\{A_m, \mathcal{O}_m\}$ by a hidden Markov model. The extension of graph based methods to other classifiers including discriminative ones can be done by introducing some function $\phi(\mathcal{O}_m | A_m)$ capable of defining for $\mathcal{O}_m$ confidence measures of belonging to different activities. Given a sequence of observations $\{\mathcal{O}_m\}_{1 \leq m \leq M}$ of length M, the sequence of activities can thus be estimated as the following :

$$\hat{A}_{1:M} = \arg\max_{A_{1:M}} p(A_1) \prod_{m=2}^{M} p(A_m | A_{m-1}) \prod_{m=1}^{M} \phi(O_m | A_m)$$
(1)

This optimization problem is solved using Viterbi algorithm. In this study no activity transitions are made impossible and the graph is used to model the temporal stable nature of activities. Next section will discuss the estimation of confidence measures based on posterior probabilities of SVM's outputs.

3. SUPPORT VECTOR MACHINES SVM

In supervised classification, we have a training set of input space $\chi = \{x_i\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^d$ $(d \geq 1)$, along with the different class tags $\{t_i\}_{i=1}^{N}$, where $t_i \in \{-1, 1\}$ $\forall i \in \{1, 2, \ldots, N\}$. The classification strategy of SVM attempts to separate the data belonging to different groups by finding the maximum margin hyperplane in the feature space. One of its advantages is its capability to deal with non-linearly separable data by using a non linear function $\psi(x)$ to map the original features into a higher dimensional space where a linear separation becomes possible. For new observed data point x, SVM make predictions based on the following function :

$$J(x) = w^T . \psi(x) + w_0 = \sum_{i \in SV} t_i \alpha_i k(x_i, x) + w_0$$
(2)

where $k(u, v) = \psi(u).\psi(v)$ is the kernel chosen *a priori*, $w = \sum_{i \in SV} t_i \alpha_i \psi(x_i)$ is the normal to the hyperplane defined in the kernel space, w_0 is the offset of the hyperplane from the origin of the space and SV is the set containing the support vectors. For a new data point x_t, the classifying decision is illustrated by a Heaviside function. The classifier chooses class 1 if $J(x_t) > 0$ and class 2 if $J(x_t) < 0$. In the case where $J(x_t) = 0$, a class is chosen at random.

3.1. Confidence measures for binary classification

The output of an SVM is not a probabilistic value but an uncalibrated distance measure of a data point to the separating hyperplane in the feature space. Nevertheless, [10] suggested one method of mapping, by means of a parametric sigmoid model, the output $J(x)$ of an SVM classifier to class probabilities $p(c = 1|x)$ and $p(c = 2|x) = 1 - p(c = 1|x)$; where class 1 is attributed positive tags and class 2 negative ones. These posterior probabilities define confidence measurements for the adherence of a new data point to the two distinct classes. Class 1 confidence measure is thus given by :

$$\phi(x|c = 1) = p(c = 1|x) = \frac{1}{1 + \exp(AJ(x) + B)}$$
(3)

The parameters A and B can be determined by minimizing the cross entropy error function of the training data. Another method of producing confidence measurements from SVM classifiers was proposed by [9]. The author used a tangent hyperbolic function, $p(c = 1|x) = \frac{1}{2}(1 + \tanh(J(x)))$. This function is a non-parametric form of equation (3) where A and B are given fixed values of -2 and 0 respectively.

The sigmoid modeling of posterior probabilities is motivated by the assumption stating that the conditional densities of the SVM outputs are exponentially distributed. However, by inspecting the inertial data of our study, the class conditional densities were more likely to follow normal distributions.

In this case, the probability density function of the output of a binary SVM is considered a sum of 2 Gaussian laws each corresponding to a class and having a mean $\mu_k(k = 1, 2)$ and a standard deviation σ_k. By defining the parameter vector $\theta = (\lambda_1, \mu_1, \sigma_1, \mu_2, \sigma_2)$, Bayes' rule allows us to write class probabilities :

$$p(c = i|x, \theta) = \frac{\lambda_i w(J(x), \mu_i, \sigma_i)}{\sum_{k=1,2} \lambda_k w(J(x), \mu_k, \sigma_k)} \quad (4)$$

Where $i \in \{1, 2\}$, $w(J(x), \mu_i, \sigma_i)$ is the Gaussian law associated to class i, λ_i represents the different proportions of classes ($\lambda_2 = 1 - \lambda_1$). The estimation of the parameter vector θ can be done by minimizing the negative conditional log-likelihood function of the training data :

$$\hat{\theta} = \arg\min_{\theta} - \sum_l \sum_{i=1,2} \mathbb{I}(x_l \in i) \log[p(c = i|x_l, \theta)] \quad (5)$$

Where $\{x_l\}_l$ is the training data and $\mathbb{I}(x)$ is an indicator function which is equal to 1 if x is true and 0 otherwise.

3.2. Confidence measures for the multi-class problem

Multi-class problems using SVMs are usually approached using a one-versus-one technique applied with a max-wins voting strategy. This method establishes one binary classifier for every possible pair of distinct classes, thus $N(N-1)/2$ binary classifiers are constructed for an N-class problem. The binary classifier $J_{i,j}$ is trained by considering positive tags for the data points belonging to class i and negative tags for those coming from class j. After presenting a new data point x to the binary classifier $J_{i,j}$, the number of votes of the class chosen by this classifier is incremented by one. When all binary classifiers make their votes, x is assigned membership to the class k with the largest number of votes (voting decision rule), $k = \arg\max_i \sum_{j,j\neq i} \mathbb{I}(\text{sgn}(J_{i,j}) > 0)$. In this case, a straightforward and simple confidence measures estimates can be derived using the Heaviside binary confidence functions, $\hat{p}_i = 2/(N(N-1)) \sum_{j,j\neq i} \mathbb{I}(\text{sgn}(J_{i,j}) > 0)$.

Another method of calculating multiclass posterior probabilities using binary posteriors was proposed by Hasite and Tibshirani [11]. The method combines the probabilistic outputs of all the binary classifiers, $p(i|i \text{ or } j, x)$ defined $\forall(i, j) \in \{1, 2, ..., N\}^2$ and $i \neq j$, to obtain estimates of posterior probabilities for all N classes, $p(i|x)$ for $i \in \{1, 2, ..., N\}$. For a data point x, let us express the probabilistic output of the binary classifier $J_{i,j}$ as $r_{i,j}(x) = p(i|i \text{ or } j, x)$. In order to evaluate estimates of the $p(i|x)$'s, $N(N-1)/2$ auxiliary variables are introduced :

$$\mu_{i,j}(x) = \frac{p(i|x)}{p(i|x) + p(j|x)} \quad (6)$$

We note that $\mu_{j,i}(x) = 1 - \mu_{i,j}(x)$ as well as $r_{j,i}(x) = 1 - r_{i,j}(x)$. It is impossible to estimate posterior probabilities by setting $\mu_{j,i}(x) = r_{j,i}(x)$ for all i,j because there

would be $N - 1$ independent parameters but $N(N - 1)/2$ independent equations. The vector of posterior probabilities $\tilde{\mathbf{p}}(\mathbf{x}) = (p(1|x), p(2, x), \ldots, p(N|x))$ is then determined so that the $\mu_{i,j}(x)$'s are sufficiently close to the $r_{i,j}(x)$'s which is done by minimizing their Kullback-Leibler distance [11].

3.2.1. Proposed method (PM)

For the estimation of the posterior probabilities $p(i|x)$'s, we propose another method that seeks the convergence of $\mu_{i,j}(x)$'s towards the $r_{i,j}(x)$'s based on the Euclidean distance. The following function $\Phi(\tilde{\mathbf{p}})$ is therefore defined :

$$\Phi(\tilde{\mathbf{p}}) = \sum_{j>i} |\mu_{i,j}(x) - r_{i,j}(x)|^2$$

The solution to this problem is based on the following theorem :

Theorem 3.1 *If $\forall i \neq j$, $r_{i,j} \in [\epsilon, 1 - \epsilon]$ with $0 < \epsilon < 0.5$, the minimal value of $\Phi(\tilde{\mathbf{p}})$ is attained for the following vector :*

$$\tilde{\mathbf{p}} = (\mathbf{1_N}(\mathbf{\Omega^T\Omega})^{-1}\mathbf{1_N^T})^{-1}(\mathbf{\Omega^T\Omega})^{-1}\mathbf{1_N^T}.$$

Where $\mathbf{\Omega}$ is a matrix of size $N(N - 1)/2 \times N$ whose coefficients are zeros except for the following ones : $\forall i \in \{1, \cdots, N - 1\}$, $\forall j \in \{i + 1, \cdots, N\}$,

$$\mathbf{\Omega}_{(i-1)/2(2N-i)+j-i,i} = r_{j,i}$$

$$\mathbf{\Omega}_{(i-1)/2(2N-i)+j-i,j} = -r_{i,j}$$

$\mathbf{1_N}$ *is the row vector of size N whose elements are equal to* 1.

The complete proof is not given here due to the lack of space but the main idea consists of addressing the optimization of $\Phi(\tilde{\mathbf{p}})$ using the KKT approach in order to take into account both constraints $\sum_i p(i|x) = 1$ and $\forall i, p(i|x) \geq 0$. The conditions $r_{i,j} \in [\epsilon; 1 - \epsilon]$ $\forall i \neq j$ ensure that $\forall i \in \{1, ..., N\}$ $p(i|x) > 0$, the thing which when combined with the KKT conditions lead to the above optimal solution.

3.2.2. Simulation : Methods comparison

Hastie and Tibshirani conducted a quick empirical evaluation in order to prove that their method outperforms the voting strategy [11]. However, in their experiment, they considered unbiased probability distribution where all posterior probabilities $p(i|x)$'s were chosen to be moderately close with a slightly higher probability for class 1. To compare the proposed method (PM) with that of Hastie and Tibshirani (HT), we conducted for 500 realizations of the added noise the same experiment, while considering this time the case where the generated a posterior probabilities can be biased. After estimating the a posterior probabilities, the winning class is chosen based on maximum a posteriori decision rule. Detection accuracy of the correct class (class 1) using different methods are

		Unbiased			Biased		
		Voting	HT	PM	Voting	HT	PM
n	3	0.92	0.95	0.97	0.98	0.99	0.99
	4	0.87	0.93	0.93	0.92	0.95	0.98
	5	0.81	0.93	0.94	0.90	0.93	0.97
	6	0.80	0.92	0.92	0.86	0.90	0.95
	8	0.79	0.94	0.94	0.82	0.88	0.94
	10	0.82	0.94	0.96	0.82	0.89	0.95

Table 1. Correct detection rate in function of the number of classes

reported in table 1. It is shown that in the general case where the distribution of probabilities is unknown (biased or not), the detection of the correct class by the PM method is better than its alternatives. This reflects the method's improved quality for providing reliable estimates of confidence measures.

4. DATA COLLECTION

Activity data was collected using a Motion POD^{TM} (MO-VEA) with a built-in 3-axial accelerometer sensor. The purpose was to classify 6 activities including lying down (LD), slouching (SL), sitting (SI), standing (ST), walking (WA) and running (RU) which principally constitute the main performed activities during daily life. The monitor sensor was placed on the belt and was used to collect, at a rate of 100 Hz, the acceleration data of 21 subjects free of chronic diseases and comprising 13 men and 8 women in the age range of 19 to 55 years (mean $\pm$ standard deviation = 36 ± 13.8). For each subject, about 2.5 hours were registered during which they carried out physical activities of varying intensities as normally as they do in their everyday life. During the experiments, the supervising medical team was charged with annotating the main performed activities of each subject.

So as to identify the ground truth, the acceleration data was visually examined with reference to the corresponding annotation records in order to define the different activity classes. Some considerations were made to huddle certain activities which are close in structure. For instance, sitting without any movement, sitting playing with a ball and sitting working on a computer were categorized as sitting ; standing without movement and standing while arms are moving were identified as standing, and finally slow and moderate intensity walking and walking on stairs were classified as walking.

5. RESULTS AND DISCUSSION

The authors in [6] examined several features and time windows and found that relatively small windows with simple features including standard deviation can yield significant recognition rates . In this study, a sliding window of 0.4s with 50 % overlapping ratio was used. The center values of the 3 accelerometer signals were retained along with the norm of the vector containing the standard deviations evaluated over

windows. To evaluate the activity classification accuracies based on the SVM classifier, a leave-one-subject-out cross validation was applied. For each subject, the binary one-versus-one SVM classifiers were trained using the radial basis function (RBF) kernel. The kernel's hyper-parameter σ and the SVM regularization parameter C were determined using 3-fold cross validation over the database. Classification performances were evaluated on the one hand for activity attribution decisions when taken by the conventional SVM method and on the other hand when decisions are reinforced by the prior knowledge modeled by the graph (equation (1)). This evaluation was repeated using the different methods estimating the binary confidence measures (Heaviside, Tanh, Sigmoid and Gaussian). Multi-class confidence measures were obtained by applying the proposed PM method due to its superior performance compared to its discussed variants. Fig. 1 displays the classification accuracies with and without graph in function of the used method for binary confidence measures estimation. Overall detection accuracy without graph was around 81% for all the 4 methods with a slight noticed advantage for the use of Gaussian modeling for the estimation of binary posterior probabilities and for which the detection rate scored 82%. This quasi tie between the methods' detection performances when no graph is used is not shocking since in this case the confidence posterior probability measures are only used to predict independently the winning activity class without attaching any importance to the quality of their estimates. In other words, as long as the true activity class for an observation vector at instant m is the one having the highest estimated confidence measure, the bias of this confidence estimate from its most representative value will not have any impact on the detected activity. We recall that without graph a good or bad detection for an observation vector at instant m doesn't affect in any way the detected activities for the neighboring observation vectors. Fig. 1 shows that the use of graph for the inclusion of prior knowledge concerning the physical activity sequences has a positive effect on the detection performances whatever the used method. Compared with the case where no graph is used, the overall detection improvement ranges between 4% to almost 10% depending on the adopted confidence estimation method. Tangent hyperbolic and Heaviside approaches for confidence estimation provide lower detection accuracies compared to the two other methods which is explained by the fact that these two approaches are non-optimal in the sense that they don't take into account the distribution of the SVM classifier outputs. The biased estimates of confidence measures should be responsible for this relative lower performance. Furthermore, using the Heaviside function a true activity class may be inconveniently attributed a confidence value equal to 0 resulting from a miss-classification by the SVM base classifier. Since the Viterbi algorithm avoids paths with null probabilities, this kind of misdetection can not be adjusted by using the graph. Gaussian modeling of confidence measures outperformed the other methods where its
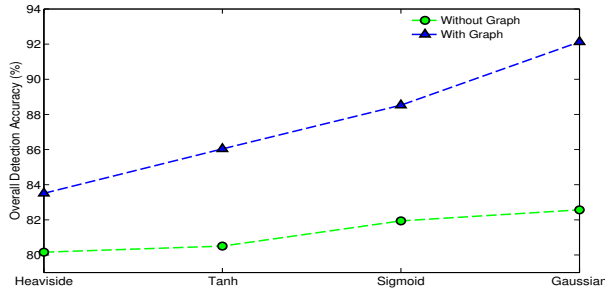
Fig. 1. Overall detection accuracies with and without graph traced in function of methods used for confidence estimation.

		Recognized Activity					
		LD	SL	SI	ST	WA	RU
Labeled Activity	LD	**0.95**	0.05	0	0	0	0
	SL	0.05	**0.93**	0.02	0	0	0
	SI	0	0.02	**0.91**	0.07	0	0
	ST	0	0	0.14	**0.86**	0	0
	WA	0	0	0	0	**0.96**	0.04
	RU	0	0	0	0	0.08	**0.92**

Table 2. Confusion matrix for SVM+Gaussian graph method

application with the graph presented a 92% recognition rate. This confirms the justness of the assumption considering normally distributed class densities for the binary SVM outputs. For the case in which the graph is applied with Gaussian modeling of confidence measures, the aggregate confusion matrix showing individual recognition rates of activities is given in Table 2. Lying down and walking activities attained relative high detection accuracies which may be ascribed to their peculiarity in acceleration data generation among other activities. Lying down produces acceleration signals whose, in general, standard deviations have values close to 0, whereas walking being a dynamic activity engenders signals having oscillating frequencies which is pointed out by larger values of standard deviation. The most significant confusion turned out to be between the sitting down and standing activities. This, however, can be justified by the occasional unsteady acceleration discriminative capacity between these two activities when the sensor is worn on belt. On the other hand, running is confused in 8% of the cases with walking which can be an interpretation of certain common characteristics between the corresponding acceleration signals, as well as the fact that running can be subject-dependent. For instance, a subject who runs less dynamically than others may have a part of his running phase detected as moderate walking which was categorized under the walking activity class during database creation.

To conclude the paper, physical activity assessment using single accelerometer worn on belt was investigated. Prior knowledge on the sequential correlation of activities was exploited in combination with confidence measures in activity classes to refine the SVM's classifier conventional decisions. For this, graph based methods were introduced and notions of statistical decision were reviewed with the purpose of providing an insight on determining and using confidence measures for physical activity classification using SVM. We then proposed a method of estimation of confidence measures using binary outputs of SVM and suggested an approach to efficiently evaluate multi-class confidence measures using binary ones. Main results showed that the best accuracy was obtained when proposed methods were applied with graph. Compared to traditional SVM classifier's decisions, results also showed that reinforcing decisions by prior knowledge increased significantly the detection rate of different activities.

6. REFERENCES

[1] U.S. Departement of Health and Human Services, "Physical activity and health : A report of the surgeon general.," *Atlanta, Georgia : US Departement of Health and Human Services, Public Health Service, CDC, National Center for Chronic Disease Prevention and Health Promotion*, pp. 3–8, 1996.

[2] Blackburn G. Higgins M. Lauer R. Perri M.G. Ryan D. Grundy, S.M., "Physical activity in the prevention and treatment of obesity and its comorbidities : Evidence report of independent panel to assess the role of physical activity in the treatment of obesity and its comorbidities," *Medicine and Science in Sports and Exercise*, vol. 31, no. 11, pp. 1493–1500, 1999.

[3] S Preece, JY Goulermas, LPJ Kenney, D Howard, K Meijer, and R Crompton, "Activity identification using body-mounted sensors a review of classification techniques," *Physiological Measurement*, vol. 30, no. 4, pp. R1, 2009.

[4] S Preece, JY Goulermas, LPJ Kenney, and D Howard, "A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data," *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 871–879, 2009.

[5] K. Altun, B. Barshan, and O. Tunel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognition*, vol. 43, no. 10, pp. 3605–3620, Oct. 2010.

[6] S. Pirttikangas, K. Fujinami, and T. Nakajima, "Feature selection and activity recognition from wearable sensors," in *Ubiquitous Computing Systems*, vol. 4239, pp. 516–527. Springer Berlin / Heidelberg, 2006.

[7] A. Y. Ng and M. I. Jordan, "On discriminative vs. generative classifiers : A comparison of logistic regression and naive bayes," 2001.

[8] P. Jallon, "A graph based algorithm for postures estimation based on accelerometers data," in *IEEE EMBC*. Aug. 2010, pp. 2778–2781, IEEE.

[9] P. Jallon, B. Dupre, and M. Antonakios, "A graph based method for timed up & go test qualification using inertial sensors," in *IEEE ICASSP*. May 2011, pp. 689–692, IEEE.

[10] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *ADVANCES IN LARGE MARGIN CLASSIFIERS*. 1999, pp. 61–74, MIT Press.

[11] Tibshirani R. Hastie, T., "Classification by pairwise coupling," *Advances in Neural Information Processing Systems*, vol. 10, pp. 507–513, 1998.