DISCRIMINATIVE SPECTRAL LEARNING OF HIDDEN MARKOV MODELS FOR HUMAN ACTIVITY RECOGNITION

Alfredo Nazábal, Antonio Artés-Rodríguez

Department of Signal and Communications Theory, Universidad Carlos III de Madrid,
Avenida de la Universidad, 30, Leganés, 28911, Spain
Email: {anazabal, antonio}@tsc.uc3m.es

ABSTRACT

Hidden Markov Models (HMMs) are one of the most important techniques to model and classify sequential data. Maximum Likelihood (ML) and (parametric and non-parametric) Bayesian estimation of the HMM parameters suffers from local maxima and in massive datasets they can be specially time consuming. In this paper, we extend the spectral learning of HMMs, a moment matching learning technique free from local maxima, to discriminative HMMs. The resulting method provides the posterior probabilities of the classes without explicitly determining the HMM parameters, and is able to deal with missing labels. We apply the method to Human Activity Recognition (HAR) using two different types of sensors: portable inertial sensors, and fixed, wireless binary sensor networks. Our algorithm outperforms the standard discriminative HMM learning in both complexity and accuracy.

***Index Terms*—** Hidden Markov Models, Spectral algorithm, Discriminative learning, Observable operator models, Human activity recognition

1. INTRODUCTION

Hidden Markov Models (HMMs) have been applied to a wide range of sequence modelling problems like speech recognition, Human Activity Recognition (HAR) or time series analysis [1]. The Expectation-Maximization (EM) algorithm is the classical method used to learn the parameters of HMMs [2]. However, it exhibits two main problems: 1) the likelihood is multimodal so the EM is guaranteed to converge only to a local maxima, and 2) although the complexity of the algorithm growths linearly with the length of the training sequence, the multiple initializations required for minimizing the effects of the local convergence and the more than quadratic growth with the number of hidden states makes EM computationally heavy with large training datasets. Bayesian inference methods including Gibbs sampling [3], variational optimization [4], or Bayesian non-parametric methods [5] are even computationally heavier and global convergence is still not guaranteed.

The authors in [6] propose a spectral algorithm for learning HMMs with discrete observations. Basically, the method adjusts the model by moment matching instead of maximizing the likelihood, and it relies on the use of the observable operators view of the HMM [7]. They use this approach to solve the prediction and filtering problems. Although the authors focus on HMMs with discrete observations, there exists several extensions for continuous observations using kernels [8, 9].

The application of HMMs to the HAR problem follows two main approaches. The first one [10] consists in learning a unique HMM, modelling only the temporal dependencies between different classes and assigning the same number of hidden states and classes. This is a very simple model and usually it must be combined with supervised learning algorithms. The second one [11] consists in learning one HMM for each possible class and then choosing the model with the maximum likelihood for each test case. The main problem of this approach is the need of defining a sequence size to learn each model and to infer the test sequences. A similar approach has been used in speech recognition where they initially train a different HMM for each class (phoneme or word) using the EM, and then they fine tune them discriminatively [12]. A more direct approach is based on dicriminative training of the HMM, computing directly the posterior probabilities of the classes [13].

In this work, we propose a spectral algorithm for learning a discriminative HMM with discrete observations. We extend the work in [6] obtaining a recursive algorithm for estimating the labels of an observation sequence. We use this approach in two different HAR settings and we compare our results with the discriminative HMM trained with the EM algorithm. We will show than under the same conditions, our algorithm outperforms the EM algorithm in computation time increasing the accuracy performance. When comparing with continuous observations, the accuracy error is slightly worse, but again, the time computation of our algorithm dramatically outperforms the EM.

This work has been partly supported by Ministerio de Economía of Spain (projects 'COMONSENS', id. CSD2008-00010, 'ALCIT', id. TEC2012-38800-C03-01, and 'COMPREHENSION', id. TEC2012-38883-C02-01).

2. SPECTRAL ALGORITHM FOR LEARNING HMMS

In this section, we briefly explain the spectral algorithm for learning HMMs with discrete observations presented in [6]. Let $Y \in \{1, \cdots, M\}$ be the hidden states of a HMM and $X \in \{1, \cdots, N\}$ the discrete observations. The probability transition matrix is $\mathbf{T} \in \mathbb{R}^{M\times M}$ where $T_{ij} = p(y_{t+1} = i|y_t = j)$, the observation probability matrix is $\mathbf{O} \in \mathbb{R}^{N\times M}$ where $O_{ij} = p(x_t = i|y_t = j)$, the initial probabilities distribution is $\boldsymbol{\pi} \in \mathbb{R}^M$ with $\pi_i = p(y_1 = i)$ and t is any time instant. We can compute the probability of a sequence of observations in terms of observable operators [7] for each observation in X:

$$\begin{aligned}\mathbf{A}_x &= \mathbf{T} \cdot \text{diag}\left(\mathbf{O}_{x,:}\right)\\ \mathbf{O}_{x,:} &= \left(O_{x,1}, \cdots, O_{x,M}\right)\end{aligned}$$

where diag$(\mathbf{O}_{x,:})$ is a diagonal matrix with elements $\mathbf{O}_{x,:}$.
For any t, the probability of any sequence of observations $x_{1:t} = [x_1, \cdots, x_t]$ can be written as the following product of matrices:

$$P\left(x_{1:t}\right) = \mathbf{1}_M^T\mathbf{A}_{x_t}\cdots\mathbf{A}_{x_1}\boldsymbol{\pi} = \mathbf{1}_M^T\mathbf{A}_{x_{t:1}}\boldsymbol{\pi} \tag{1}$$

where $\mathbf{1}_M$ is a column vector of M ones. This expression depends on the transition matrix $\mathbf{T}$ and the observation matrix $\mathbf{O}$ of the model, but neither are known at the training stage.
If we are only interested in the value of (1) and not in the parameters of the model, we can use an invertible transformation of this equation that only depends on observable quantities:

$$\begin{aligned}\mathbf{1}_M^T\mathbf{A}_{x_{t:1}}\boldsymbol{\pi} &= \mathbf{1}_M^T\mathbf{S}^{-1}\mathbf{S}\mathbf{A}_{x_t}\mathbf{S}^{-1}\cdots\mathbf{S}\mathbf{A}_{x_1}\mathbf{S}^{-1}\mathbf{S}\boldsymbol{\pi}\\ &= \mathbf{b}_\infty^T\mathbf{B}_{x_{t:1}}\mathbf{b}_1\end{aligned} \tag{2}$$

where $\mathbf{S} \in \mathbb{R}^{M\times M}$ is the invertible transformation matrix and

$$\mathbf{b}_\infty^T = \mathbf{1}_M^T\mathbf{S}^{-1}, \quad \mathbf{B}_x = \mathbf{S}\mathbf{A}_x\mathbf{S}^{-1}, \quad \mathbf{b}_1 = \mathbf{S}\boldsymbol{\pi}$$

The idea of this transformation is to avoid the identification problem of the hidden structure of the model by expressing (2) in terms of the marginal probabilities of the vector of singletons $\mathbf{p}_1 \in \mathbb{R}^N$, matrix of pairs $\mathbf{P}_{21} \in \mathbb{R}^{N\times N}$ and tensor of triples $\underline{\mathbf{P}_{31}^x} \in \mathbb{R}^{N\times N}, \forall x$. These quantities can be estimated from the data, and they are related to the parameters of the HMM

$$\begin{aligned}\mathbf{p}_1 &= P(x_1) = \mathbf{O}\boldsymbol{\pi}\\ \mathbf{P}_{21} &= P(x_2, x_1) = \mathbf{O}\mathbf{T}\text{diag}\left(\boldsymbol{\pi}\right)\mathbf{O}^T\\ \underline{\mathbf{P}_{31}^x} &= P(x_3, x_2 = x, x_1) =\\ &= \mathbf{O}\mathbf{A}_x\mathbf{T}\text{diag}\left(\boldsymbol{\pi}\right)\mathbf{O}^T, \forall x\end{aligned}$$

In [6], the authors proof that $\mathbf{S} = \mathbf{U}^T\mathbf{O}$ is a valid invertible transformation for the HMM model, where $\mathbf{U} \in \mathbb{R}^{N\times M}$ is the matrix of M left singular vectors of the joint probability matrix $\mathbf{P}_{21}$. With this transformation, we can express the new parameters of the model as:

$$\begin{aligned}\mathbf{b}_1 &= \mathbf{U}^T\mathbf{p}_1\\ \mathbf{b}_\infty &= \left(\mathbf{P}_{21}^T\mathbf{U}\right)^\dagger\mathbf{p}_1\\ \mathbf{B}_x &= \left(\mathbf{U}^T\underline{\mathbf{P}_{31}^x}\right)\left(\mathbf{U}^T\mathbf{P}_{21}\right)^\dagger, \forall x\end{aligned} \tag{3}$$

where $(\cdot)^\dagger$ is the Moore-Penrose pseudo-inverse operation.

To compute the spectral algorithm for learning a HMM, first we take m *i.i.d.* triples $[x_1, x_2, x_3]$ from the training observations to obtain an estimation of $\mathbf{p}_1$, $\mathbf{P}_{21}$ and $\underline{\mathbf{P}_{31}^x}$. Then, we compute the SVD of $\mathbf{P}_{21}$, obtaining the matrix of M left singular vectors $\mathbf{U}$. Finally, we compute the HMM transformed parameters in (3).

3. SPECTRAL ALGORITHM FOR LEARNING DISCRIMINATIVE HMMS

The algorithm in Section 2 can be used to obtain the probability of a sequence of observations or the probability of the last observation given all the previous ones. However, the focus of this work is the classification of a sequence of discrete labels given the observations, using the discriminative HMM model in Figure 1. We want to solve 1) the problem of predicting the probability of a sequence of labels given the observations and 2) the problem of predicting the conditional probability of a label given all the previous labels and observations.

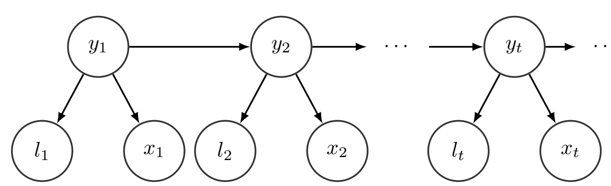

Fig. 1. Discriminative HMM graphical model of a sequence of t observations $x_{1:t}$ and labels $\ell_{1:t}$

Let $L \in \{1, \cdots, K\}$ be the alphabet of labels of the model and $\mathbf{D} \in \mathbb{R}^{K\times M}$ the label probability matrix where $D_{ij} = p(\ell_t = i|y_t = j)$. We define the joint probability of a sequence of labels $\ell_{1:t}$ and observations $x_{1:t}$ as in (1)

$$P(\ell_{1:t}, x_{1:t}) = \mathbf{1}_M^T\mathbf{A}_{\ell_t x_t}\cdots\mathbf{A}_{\ell_1 x_1}\boldsymbol{\pi} \tag{4}$$

where the label and observation operators are

$$\mathbf{A}_{\ell_t x_t} = \mathbf{T}\text{diag}(\mathbf{D}_{\ell_t,:})\text{diag}(\mathbf{O}_{x_t,:}) \tag{5}$$

We have NK label and observation operators, the combination of all its possible values. In this discriminative HMM model, the labels and the observations are independent given the hidden states, so we can define for simplicity a joint label-observation matrix $\mathbf{F} \in \mathbb{R}^{NK\times M}$ where

$$F_{ijk} \quad = \quad P(\ell_t = i|y_t = k)P(x_t = j|y_t = k) = D_{ik}O_{jk}$$

and from (5), our new operators are $\mathbf{A}_{\ell_t x_t} = \mathbf{T}\text{diag}(\mathbf{F}_{\ell_t x_t,:})$.

To solve the first problem, we compute the probability of a sequence of labels given the observations applying the Bayes Theorem and combining the expressions (1) for the denominator and (4) for the numerator:

$$P(\ell_{1:t}|x_{1:t}) = \frac{P(\ell_{1:t}, x_{1:t})}{P(x_{1:t})} = \frac{\mathbf{1}_M^T \mathbf{A}_{\ell_t x_t} \cdots \mathbf{A}_{\ell_1 x_1} \boldsymbol{\pi}}{\mathbf{1}_M^T \mathbf{A}_{x_t} \cdots \mathbf{A}_{x_1} \boldsymbol{\pi}} \quad (6)$$

We are again only interested in the value of (6), and not in the values of the matrices $\mathbf{T}$, $\mathbf{O}$ and $\mathbf{D}$. We can apply an invertible transformation in the numerator and the denominator of this expression so it only depends on the training observations and labels. We follow Section 2 to obtain the denominator, and we use an invertible transformation matrix $\mathbf{R} \in \mathbb{R}^{NK \times M}$ as in (2) to obtain the numerator:

$$\mathbf{1}_M^T \mathbf{A}_{\ell_{t:1} x_{t:1}} \boldsymbol{\pi} = \mathbf{c}_\infty^T \mathbf{C}_{\ell_{t:1} x_{t:1}} \mathbf{c}_1$$

where $\mathbf{A}_{\ell_{t:1} x_{t:1}}$ is the product $\mathbf{A}_{\ell_t x_t} \cdots \mathbf{A}_{\ell_1 x_1}$ and

$$\mathbf{c}_\infty^T = \mathbf{1}_M^T \mathbf{R}^{-1}, \quad \mathbf{C}_{\ell x} = \mathbf{R} \mathbf{A}_{\ell x} \mathbf{R}^{-1}, \quad \mathbf{c}_1 = \mathbf{R}\boldsymbol{\pi} \quad (7)$$

We want to represent the numerator in (6) in terms of the observable vector $\mathbf{q}_1 \in \mathbb{R}^{NK}$, matrix $\mathbf{Q}_{21} \in \mathbb{R}^{NK \times NK}$ and tensor $\underline{\mathbf{Q}_{31}^{\ell x}} \in \mathbb{R}^{NK \times NK}, \forall \ell, x$. Also, these observable quantities can be expressed in terms of the hidden parameters of the model.

$$\begin{aligned} \mathbf{q}_1 &= P(\ell_1, x_1) = \mathbf{F}\boldsymbol{\pi} \\ \mathbf{Q}_{21} &= P(\ell_2, x_2, \ell_1, x_1) = \mathbf{F}\mathbf{T}\text{diag}\,(\boldsymbol{\pi})\,\mathbf{F}^T \\ \underline{\mathbf{Q}_{31}^{\ell x}} &= P(\ell_3, x_3, \ell_2 = \ell, x_2 = x, \ell_1, x_1) = \\ &= \mathbf{F}\mathbf{A}_{\ell x}\mathbf{T}\text{diag}\,(\boldsymbol{\pi})\,\mathbf{F}^T \end{aligned}$$

From [6] it is straightforward to proof that $\mathbf{R} = \mathbf{V}^T\mathbf{F}$ is a valid invertible transformation matrix for the model, where $\mathbf{V}$ is the matrix of left singular vectors of $\mathbf{Q}_{21}$. We can calculate the transformed parameters of the model in (7) in terms of $\mathbf{V}$ and these observable quantities:

$$\begin{aligned} \mathbf{c}_1 &= \mathbf{V}^T \mathbf{q}_1 \\ \mathbf{c}_\infty &= (\mathbf{Q}_{21}^T \mathbf{V})^\dagger \mathbf{q}_1 \\ \mathbf{C}_{lx} &= (\mathbf{V}^T \underline{\mathbf{Q}_{31}^{\ell x}})(\mathbf{V}^T \mathbf{Q}_{21})^\dagger, \forall \ell, x \end{aligned}$$

To solve the second problem, we use the Bayes Theorem to express the conditional probability of a label ℓ_t given the all the previous labels $\ell_{1:t-1}$ and all the observations $x_{1:t}$

$$P(\ell_t|\ell_{1:t-1}, x_{1:t}) = \frac{\mathbf{c}_\infty^T \mathbf{C}_{\ell_{t:1} x_{t:1}} \mathbf{c}_1}{\mathbf{b}_\infty^T \mathbf{B}_{x_t} \mathbf{Z} \mathbf{C}_{\ell_{t-1:1} x_{t-1:1}} \mathbf{c}_1} \quad (8)$$

where $\mathbf{Z} = (\mathbf{U}^T\mathbf{O})(\mathbf{V}^T\mathbf{F})$ is the transformation term between $\mathbf{B}_{x_t}$ and $\mathbf{C}_{\ell_{t-1} x_{t-1}}$. We can further express $\mathbf{Z}$ in terms of an additional observable operator $\mathbf{W}_{21} \in \mathbb{R}^{N \times NK}$

$$\begin{aligned} \mathbf{W}_{21} &= P(x_2, \ell_1, x_1) = \mathbf{O}\mathbf{T}\text{diag}\,(\boldsymbol{\pi})\,\mathbf{F}^T \\ \mathbf{Z} &= (\mathbf{U}^T \mathbf{W}_{21})(\mathbf{V}^T \mathbf{Q}_{21})^\dagger \end{aligned}$$

Finally, equation (8) can also be represented recursively as follows:

$$\begin{aligned} P(\ell_t|\ell_{1:t-1}, x_{1:t}) &= \frac{\mathbf{c}_\infty^T \mathbf{C}_{\ell_t x_t} \mathbf{c}_t}{\sum_\ell \mathbf{c}_\infty^T \mathbf{C}_{\ell x_t} \mathbf{c}_t} \\ \mathbf{c}_{t+1} &= \frac{\mathbf{C}_{\ell_t x_t} \mathbf{c}_t}{\mathbf{b}_\infty^T \mathbf{B}_{x_{t+1}} \mathbf{Z} \mathbf{C}_{\ell_t x_t} \mathbf{c}_t} \end{aligned} \quad (9)$$

We describe the spectral algorithm to solve both classification problems for discriminative HMMs in Algorithm 1. It is interesting to notice that this algorithm can also handle missing labels in the training. If we take a triple $\{x_1, x_2, x_3\}$ with missing labels, we can just actualize $\mathbf{p}_1$, $\mathbf{P}_{21}$ and $\underline{\mathbf{P}_{31}^x}$ without changing the other operators and proceed normally.

Algorithm 1 Spectral Algorithm for discriminative HMMs

Input:
Number of hidden states M and labels K.
Sample size N
m *i.i.d.* groups of triples $\{x_1, x_2, x_3\}$ and $\{\ell_1, \ell_2, \ell_3\}$,
Test sequence $x_{1:t}$

Output: $P(\ell_t|\ell_{1:t-1}, x_{1:t})$

1. Compute the empirical estimates of $\hat{\mathbf{p}}_1$, $\hat{\mathbf{P}}_{21}$, $\underline{\hat{\mathbf{P}}_{31}^x}$, $\hat{\mathbf{q}}_1$, $\hat{\mathbf{Q}}_{21}$, $\underline{\hat{\mathbf{Q}}_{31}^{\ell x}}$ and $\hat{\mathbf{W}}_{21}$.
2. Compute both the SVD of $\hat{\mathbf{P}}_{21}$ to get the matrix of M left singular vectors $\hat{\mathbf{U}}$ and the SVD of $\hat{\mathbf{Q}}_{21}$ to get the matrix of M left singular vectors $\hat{\mathbf{V}}$.
3. Compute the transformed HMM model quantities $\hat{\mathbf{b}}_1$, $\hat{\mathbf{b}}_\infty$, $\hat{\mathbf{B}}_x$, $\hat{\mathbf{c}}_1$, $\hat{\mathbf{c}}_\infty$, $\hat{\mathbf{C}}_{\ell x}$ and $\hat{\mathbf{Z}}$.
4. Compute sequentially for all t the probabilities of the labels $\ell_{1:t}$ using (9).

4. RESULTS

4.1. Inertial sensors database

In the first experiment, we use a HAR database obtained using miniature inertial sensors from APDM [14], which provide acceleration, gyroscope and magnetometer data. 18 different people were used as subjects for the experiments and only one sensor was placed at the waist of each of them. Then, they were asked to perform a sequence of activities, that were combinations of running, walking, standing, sitting and lying, in no particular order. Also, the data was processed to make it invariant to sensor orientation, following the work in [15]. The bottleneck of the spectral algorithm is the computation of the SVD. However, we can overcome this problem by substituting the SVD by a random normalized matrix and a regularization step, following the work in [16].

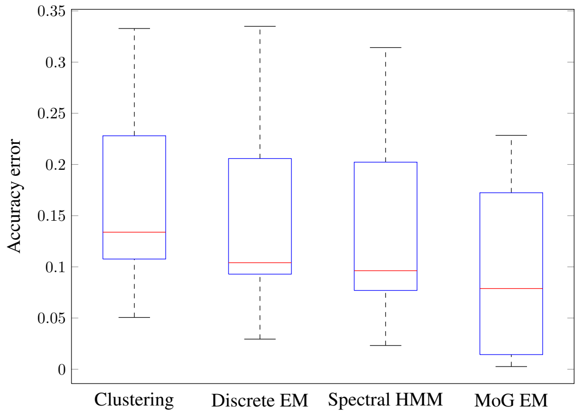

Fig. 2. Accuracy error of the activity classification using a clustering classification, the discriminative HMM with the EM algorithm, both with continuous and discrete observations, and the discriminative HMM with spectral learning and discrete observations.

To evaluate our algorithm, we get 1000 random unique observations as centroids from the training data and assign each observation to one of these centroids according to the minimum euclidean distance. We re-estimate the centroids 100 times getting the mean accuracy error for each case[1]. Also, the number of centroids for each activity is proportional to the number of training observations of each activity. We perform a leave-one-out strategy, where we use 17 sequences of activities as training and one sequence as test to evaluate the performance of the algorithms. The number of hidden states is set to $M = 25$. For the EM algorithm, we perform 5 iterations of the algorithm with k-means initialization. We consider two different EMs, one with continuous observations, assuming that the probability of the observations is a Mixture of Gaussians (MoG) and the other one with the same discrete observations as with the spectral case. We include the results from performing clustering on the observations using the minimum euclidean distance as a baseline performance.

In Figure 2 we show the comparison of all the methods for each leave-one-out case. We observe that the accuracy error for the MoG EM is in mean 3.39% better than in the spectral learning. Also, the discrete EM algorithm performs a 1.40% worse than the spectral learning algorithm. We can conclude that the reason for the loss in accuracy is due to the use of discrete observations. In fact, under the same assumptions, the absence of local minima in the spectral algorithm results in a better performance.

The most important point of this comparison is the drastic difference in computation time. The EM algorithm requires multiple initializations to avoid the local minima. Furthermore, the EM algorithm is a recursive algorithm with the number of observations, meanwhile this restriction is not present in the spectral algorithm. In particular, in a 2.5 Ghz Intel Core i5 processor with Matlab and C for the recursive section of the EM, 30.9 minutes are needed in the training step for one of the sequences of the leave-one-out case, while in the spectral algorithm case implemented exclusively in Matlab, only 1.37 seconds are needed.

[1]The standard deviation for all the cases is less than 0.04.

4.2. Binary sensors database

In the second experiment, we have used the databases with discrete observations generated by the authors in [17]. They implemented a binary sensor network in three different home environments to detect several events, like movement of objects or motion in a specific area. They placed 14, 23 and 21 sensors in each house respectively and they tagged between 10 and 16 activities. We use the last-fired feature representation of the database, where a sensor remains activated as long as another sensor is not activated. We compare the results of the HMM using the EM algorithm with the spectral algorithm for each of the three settings in terms of the accuracy error in Table 1. We can see that our spectral algorithm obtains better results in the first two houses while performing worse in the other one.

House	HMM+EM	Spectral HMM
A	0.105	0.053
B	0.516	0.332
C	0.161	0.227

Table 1. Accuracy error of the activity classification using both the discriminative HMM with the EM algorithm and the discriminative HMM with spectral learning and discrete observations.

5. CONCLUSIONS

In this work, we have proposed a spectral algorithm for learning a discriminative HMM with discrete observations. Our algorithm exhibits a similar performance as the continuous observations HMM while dramatically outperforming it in terms of time computation. Also, it performs better in general than the discrete HMM, leading to the hypothesis that the loss of accuracy is due to the discreteness of the observations, not to the spectral algorithm implementation. The low computation time and complexity and the possibility of obtaining directly the probability of the labels make it useful for on-line tracking problems, e.g. human on-line activity recognition applications. In the future, we will try to extend this spectral algorithm for discriminative HMMs with continuous observations and for other models where the large amount of training data make iterative methods prohibitive.

6. REFERENCES

[1] L. Rabiner and B.H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, 1993.

[2] L.E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," *Ann. Math. Statist.*, pp. 164–171, 1970.

[3] C. P. Robert, G. Celeux, and J. Diebolt, "Bayesian estimation of hidden markov chains: A stochastic implementation," *Statistics & Probability Letters*, vol. 16, no. 1, pp. 77–83, 1993.

[4] D.J.C. MacKay, "Ensemble learning for hidden markov models," Tech. Rep., Technical report, Cavendish Laboratory, University of Cambridge, 1997.

[5] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical dirichlet processes," *Journal of the american statistical association*, vol. 101, no. 476, 2006.

[6] D. Hsu, S. M. Kakade, and T. Zhang, "A spectral algorithm for learning hidden markov models," *Journal of Computer and System Sciences*, vol. 78, no. 5, pp. 1460–1480, 2012.

[7] H. Jaeger, "Observable operator models for discrete stochastic time series," *Neural Computation*, vol. 12, no. 6, pp. 1371–1398, 2000.

[8] S. M. Siddiqi, B. Boots, and G. J. Gordon, "Reduced-rank hidden markov models," *arXiv*, 2009.

[9] L. Song, B. Boots, S. M. Siddiqi, G. J. Gordon, and A. J. Smola, "Hilbert space embeddings of hidden markov models," in *Proceedings of the 27th international conference on machine learning (ICML)*, 2010, pp. 991–998.

[10] C. Zhu and W. Sheng, "Recognizing human daily activity using a single inertial sensor," in *8th World Congress on Intelligent Control and Automation (WCICA)*. IEEE, 2010, pp. 282–287.

[11] C. W. Han, S. J. Kang, and N. S. Kim, "Implementation of hmm-based human activity recognition using single triaxial accelerometer," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 93, no. 7, pp. 1379–1383, 2010.

[12] F. Jelinek, *Statistical methods for speech recognition*, MIT press, 1997.

[13] P. C. Woodland and D. Povey, "Large scale discriminative training of hidden markov models for speech recognition," *Computer Speech & Language*, vol. 16, no. 1, pp. 25–47, 2002.

[14] APDM Inc., "http://www.apdm.com/," .

[15] V. Elvira, A. Nazabal-Renteria, and A. Artes-Rodriguez, "A novel feature extraction technique for human activity recognition," in *IEEE Workshop on Statistical Signal Processing (SSP)*. IEEE, 2014, pp. 177–180.

[16] H. Q. Minh, M. Cristani, A. Perina, and V. Murino, "A regularized spectral algorithm for hidden markov models with applications in computer vision," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2384–2391.

[17] T.L.M. van Kasteren, G. Englebienne, and B.J.A. Kröse, "Human activity recognition from wireless sensor network data: Benchmark and software," in *Activity Recognition in Pervasive Intelligent Environments*, pp. 165–186. Springer, 2011.