A Novel Low-Complexity HMM Similarity Measure

Article *in* IEEE Signal Processing Letters · March 2011

DOI: 10.1109/LSP.2010.2096417 · Source: IEEE Xplore

2 authors, including:

Byung-Jun Yoon
Texas A&M University

116 PUBLICATIONS **971** CITATIONS

SEE PROFILE

A Novel Low-Complexity HMM Similarity Measure

Sayed Mohammad Ebrahim Sahraeian, *Student Member, IEEE,*
and Byung-Jun Yoon, *Member, IEEE*

Abstract

In this letter, we propose a novel similarity measure for comparing Hidden Markov models (HMMs) and an efficient scheme for its computation. In the proposed approach, we probabilistically evaluate the correspondence, or goodness of match, between every pair of states in the respective HMMs, based on the concept of semi-Markov random walk. We show that this correspondence score reflect the contribution of a given state pair to the overall similarity between the two HMMs. For similar HMMs, each state in one HMM is expected to have only a few matching states in the other HMM, resulting in a sparse state correspondence score matrix. This allows us to measure the similarity between HMMs by evaluating the sparsity of the state correspondence matrix. Estimation of the proposed similarity score does not require time-consuming Monte-Carlo simulations, hence it can be computed much more efficiently compared to the Kullback-Leibler divergence (KLD) thas has been widely used. We demonstrate the effectiveness of the proposed measure through several examples.

Index Terms

Hidden Markov model (HMM) similarity measure, semi-Markov random walk, HMM comparison.

EDICS Category: SAS-STAT

The authors are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128, USA (e-mail: msahraeian@tamu.edu; bjyoon@ece.tamu.edu). The correspondence may be addressed to Byung-Jun Yoon, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA. (Phone: 979-845-6942, Fax: 979-845-6259, Email: bjyoon@ece.tamu.edu).

A Novel Low-Complexity HMM Similarity Measure

I. Introduction

Due to the effectiveness in modeling symbol sequences, Hidden Markov models (HMMs) have been widely used in various applications, such as speech recognition, image classification, and computational biology. Many applications that employ HMMs for statistical data modeling require an effective similarity measure for comparing HMMs. For example, this may be necessary in problems that involve clustering or classification of HMMs, or in the parameter re-estimation process for training an HMM [1], [2]. Specific applications can be found in texture retrieval and classification [3], [4], handwriting recognition [5], protein homology detection [6], and speech recognition [7].

A conventional way for comparing two HMMs is to use the Kullback-Leibler divergence (KLD) [8]. KLD measures the distance between two probability density functions (PDFs) $P(x)$ and $\hat{P}(x)$ as:

$$D_{\mathrm{KL}}(P(x)\|\hat{P}(x)) = \int_x P(x)\log\frac{P(x)}{\hat{P}(x)}\,dx. \tag{1}$$

Since no closed-form expression exists for KLD between two HMMs, Juang and Rabiner [1] proposed a Monte-Carlo approximation of this integral for comparing two HMMs λ_1 and λ_2 using a sequence of observed symbols:

$$D_{\mathrm{KL}}(\lambda_1\|\lambda_2) \simeq \frac{1}{T}(\log P(O|\lambda_1) - \log P(O|\lambda_2)), \tag{2}$$

where $O = o_1 o_2 \cdots o_T$ is an observation sequence of length T, generated by the model λ_1. Despite its simplicity, this approximation has several limitations that make it unsuitable for some practical applications. First of all, this method has high computational cost, since T should be large enough to obtain an accurate estimate of $D_{\mathrm{KL}}(\lambda_1\|\lambda_2)$. Second of all, the estimate may vary due to the random nature of the Monte-Carlo technique. To overcome this problem, Do [2] proposed a technique for fast computation of an upper-bound of the KLD for HMMs and dependence trees, and Silva and Narayanan [7] reported an upper-bound of the KLD for transient HMMs. In addition to the KLD, several other measures have been proposed for comparing HMMs. Examples include measures based on co-emission probability of two HMMs [9], generalized probability product kernel [4], HMM stationary cumulative distribution function [10], and Bayes probability of error [5].

In this letter, we introduce a novel low-complexity HMM similarity measure that can accurately evaluate the similarity between two HMMs. The proposed measure can be efficiently computed based on the concept of semi-Markov random walk without requiring time-consuming Monte-Carlo simulations. Through several examples, we demonstrate that this measure can effectively capture the similarity between HMMs and can be computed using only a small fraction of time needed for computing KLD.

II. HMM Similarity Measure

A. Motivation and Approach

Let $\lambda = (\mathbf{p}, \mathbf{A}, \mathbf{B})$ denote an HMM with parameters $\mathbf{p}$, $\mathbf{A}$, and $\mathbf{B}$, where $\mathbf{p}$ is the initial state distribution vector, $\mathbf{A} = [a_{i,j}]$ is the state transition probability matrix ($a_{i,j}$ is the transition probability from state v_i to state v_j), and $\mathbf{B} = \{\mathbf{b}_i\}$ is the set of symbol emission probability distributions ($\mathbf{b}_i$ is the emission probability distribution at state v_i). Suppose there are two HMMs $\lambda = (\mathbf{p}, \mathbf{A}, \mathbf{B})$ and $\lambda' = (\mathbf{p}', \mathbf{A}', \mathbf{B}')$, whose similarity we want to evaluate. One way to evaluate the similarity between these HMMs is to consider the probability that the two HMMs will generate identical observations. Consider the emission of the kth symbol by the respective HMMs, whose underlying states are $s_k = v_i$ in the HMM λ and $s'_k = v'_{i'}$ in λ'. The probability that both HMMs will emit the same symbol will depend on the emission probabilities $\mathbf{b}_i$ and $\mathbf{b}'_i$, and a probabilistic similarity (or distance) measure $S_e(v_i, v'_{i'})$, such as the KLD, can be used to estimate how likely it is that the two HMMs will generate identical symbols. If we take expectation of this measure $S_e(v_i, v'_{i'})$ over all possible state pairs $(v_i, v'_{i'})$, the expected similarity between λ and λ' is given as:

$$ES(\lambda, \lambda') = \mathbf{E}[S_e(v_i, v'_{i'})] = \sum_{\forall i} \sum_{\forall i'} \pi_i \pi'_i S_e(v_i, v'_{i'}), \tag{3}$$

where π and π' are the stationary distributions of the underlying Markov chains of λ and λ', respectively, assuming that they exist. $ES(\lambda, \lambda')$ reports how likely it is that λ and λ' will emit the same symbol at a given time, based on their state transition probabilities and the emission probability distributions. Although $ES(\lambda, \lambda')$ can provide a measure of the overall similarity between the HMMs, it does not explicitly capture the similarities and differences between each pair of states $v_i \in \lambda$ and $v'_{i'} \in \lambda'$ in the respective HMMs. Instead, we can compare individual states based on how likely it is that the respective HMMs will be at v_i and $v'_{i'}$ at a given

time, and how likely they are to emit identical symbols. This leads to the definition of the *state correspondence matrix* $\mathbf{Q} = [q_{i,i'}]$, where $q_{i,i'}$ is the *correspondence score* between two states v_i in λ and $v'_{i'}$ in λ', defined as the contribution of each state pair $(v_i, v'_{i'})$ in the expected similarity $ES(\lambda, \lambda')$:

$$q_{i,i'} = \frac{\pi_i \pi'_{i'} S_e(v_i, v'_{i'})}{ES(\lambda, \lambda')} \tag{4}$$

In general, the matrix $\mathbf{Q}$ is sparse for similar HMMs, since each state in one HMM will have only a few (ideally, only one) matching states in the other HMM. This observation allows us to define a low-complexity HMM similarity measure based on the sparsity of the state correspondence matrix $\mathbf{Q}$.

B. Semi-Markov Random Walk

Interestingly, we can interpret the state correspondence score $q_{i,i'}$ in (4) based on the concept of *semi-Markov random walk*. A semi-Markov process is a stochastic process that makes state transitions according to a Markov chain but spends a random amount of time between transitions [11]. Let the graphs G and G' represent the state transition diagrams of λ and λ', respectively. Assume that we perform a simultaneous random walk on both G and G', according to the underlying state transition probabilities $\mathbf{A}$ and $\mathbf{A}'$ of the corresponding HMMs. This is equivalent to performing a random walk on the product graph $G_\times$ of G and G', where the nodes in $G_\times$ correspond to pairs of states $(v_i, v'_{i'})$ in the respective HMMs and the probability for making a transition from a node $(v_i, v'_{i'})$ in $G_\times$ to another node $(v_j, v'_{j'})$ is the product $a_{i,j} a'_{i',j'}$ of the corresponding transition probabilities in λ and λ'. The stationary distribution of the random walk on this product graph is given by $\pi_\times = \pi \otimes \pi'$, where the stationary distributions π and π' correspond to the normalized left eigenvectors of the state transition probability matrices $\mathbf{A}$ and $\mathbf{A}'$, respectively, and $\otimes$ is the Kronecker product [12]. Now assume that the walker spends an average amount of time μ_j once it enters a given node $v_{\times j} = (v_i, v'_{i'})$ in the product graph $G_\times$. If the mean time spent at the node $v_{\times j}$ is proportional to the similarity between two nodes (i.e., states) v_i and $v'_{i'}$, such that $\mu_j \propto S_e(v_i, v'_{i'})$, the proportion of time that the walker will spend at $v_{\times j} = (v_i, v'_{i'})$ will be

$$\tilde{\pi}_{\times_j} = \frac{\pi_{\times_j} \mu_j}{\sum_{\forall j} \pi_{\times_j} \mu_j} = \frac{\pi_i \pi'_{i'} S_e(v_i, v'_{i'})}{\sum_{\forall i,i'} \pi_i \pi'_{i'} S_e(v_i, v'_{i'})}. \tag{5}$$

We can see that $\tilde{\pi}_{\times_j}$ is identical to the state correspondence score $q_{i,i'}$ shown in (4). Thus, we can also view $q_{i,i'}$ as the long-run proportion of time that the random walker stays at

$v_{\times j} = (v_i, v'_{i'})$, or equivalently, the proportion of time that it stays simultaneously at v_i in G and $v'_{i'}$ in G'. According to this model, $q_{i,i'} = \tilde{\pi}_{\times_j}$ will be larger if the symbol emission probability distributions $\mathbf{b}_i$ and $\mathbf{b}'_i$ are more similar to each other. Furthermore, $q_{i,i'}$ also incorporates the stationary probability of visiting the states v_i and $v'_{i'}$, and it increases as π_i and $\pi'_{i'}$ increase. In this way, each element $q_{i,i'}$ of the state correspondence matrix $\mathbf{Q}$ provides an effective measure of similarity between a pair of states in the respective HMMs.

C. Comparing Symbol Emission Probability Distributions

We can evaluate the similarity between states as follows:

$$S_e(v_i, v'_{i'}) = 1/D(\mathbf{b}_i \| \mathbf{b}'_{i'}) \text{ or } S_e(v_i, v'_{i'}) = e^{-kD(\mathbf{b}_i \| \mathbf{b}'_{i'})}, \tag{6}$$

by using a distance (or dissimilarity) measure $D(\mathbf{b}_i \| \mathbf{b}'_{i'})$ for the underlying symbol emission probability distribution $\mathbf{b}_i$ (at state v_i) and $\mathbf{b}'_{i'}$ (at state $v'_{i'}$), where k is a constant. We have several different choices for $D(.\|.)$ [13], [14], where some examples are shown below. In what follows, we assume that the HMMs emit discrete observations for simplicity, but extension to the continuous case is straightforward.

1) Symmetric KLD: Defined as $\frac{1}{2}[D_{\mathrm{KL}}(P\|Q) + D_{\mathrm{KL}}(Q\|P)]$, where $D_{\mathrm{KL}}(P\|Q)$ is the KLD between two probability mass functions P and Q.

2) λ-divergence: An alternative to the original KLD defined as $D_\lambda(P\|Q) = \lambda D_{\mathrm{KL}}\big(P\|\lambda P + (1-\lambda)Q\big) + (1-\lambda) D_{\mathrm{KL}}\big(Q\|\lambda P + (1-\lambda)Q\big)$. For $\lambda = 0.5$, this measure reduces to the Jensen-Shannon divergence.

3) α-divergence: Defined as $D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log(\sum_i P(i)^\alpha Q(i)^{1-\alpha})$, for parameter $\alpha > 0$. This measure reduces to the Bhattacharyya distance for $\alpha = 0.5$ (except for a scale factor of two).

D. Measuring the Sparsity of the State Correspondence Matrix

As discussed earlier, we expect that the state correspondence matrix $\mathbf{Q}$ will be sparse when the two HMMs λ and λ' are similar to each other. More precisely, every row and column in $\mathbf{Q}$ should be sparse when we have similar HMMs, and in an ideal case, $\mathbf{Q}$ will resemble a permutation matrix. Therefore, we can measure the similarity between HMMs by evaluating the sparsity of the state correspondence matrix $\mathbf{Q}$. This is conceptually similar to using a graph kernel framework for comparing the topology of the given HMMs [12]. Graph kernels compare substructures of graphs (such as walks, paths, or trees) that are computable in polynomial

time to measure the similarity between two graphs. In our method, we use this concept in a probabilistic fashion, in terms of semi-Markov random walk on the product graph of the two HMMs.

Based on this idea, we propose the following similarity measure for comparing two HMMs λ and λ':

$$S(\lambda\|\lambda') \triangleq \frac{1}{2}\left[\frac{1}{M}\sum_{i=1}^{M} H(\mathbf{r}_i) + \frac{1}{M'}\sum_{j=1}^{M'} H(\mathbf{c}_j)\right], \tag{7}$$

where $\mathbf{r}_i$ is the ith row of $\mathbf{Q}$, $\mathbf{c}_j$ is the jth column of $\mathbf{Q}$, and M and M' respectively are the number of rows and columns of $\mathbf{Q}$, which correspond to the number of states in λ and λ'. The function $H(\mathbf{u})$ is a normalized sparsity measure for the vector $\mathbf{u}$. There are various methods for measuring sparsity [15], where examples include the norm measures (l^0, l^0_ϵ, l^1, l^p, $\frac{l^2}{l^1}$), entropy measures, Hoyer measure, pq-mean, and the Gini Index. In this work, we adopt the *Gini Index*, since it has many useful properties that make it a good sparsity measure [15]. The Gini Index of a vector $\mathbf{u}$ of length N is defined as:

$$GI(\mathbf{u}) = 1 - 2\sum_{k=1}^{N} \frac{u_{(k)}}{\|\mathbf{u}\|_1}\left(\frac{N-k+\frac{1}{2}}{N}\right), \tag{8}$$

where $\|\mathbf{u}\|_1$ is the ℓ_1 norm of $\mathbf{u}$, and $u_{(k)}$ is the kth smallest element of $\mathbf{u}$, such that $u_{(1)} \leq u_{(2)} \leq \cdots \leq u_{(N)}$. In (7), we use the normalized Gini Index $H(\mathbf{u})$, which is defined as:

$$H(\mathbf{u}) \triangleq \frac{N \cdot GI(\mathbf{u})}{N-1} = \frac{N}{N-1} - 2\sum_{k=1}^{N} \frac{u_{(k)}}{\|\mathbf{u}\|_1}\left(\frac{N-k+\frac{1}{2}}{N-1}\right), \tag{9}$$

for $N > 1$. For $N = 1$, we define $H(\mathbf{u}) \triangleq 0$. According to this definition, a vector $\mathbf{u_0}$ with only one nonzero element attains the maximum sparsity score of $H(\mathbf{u_0}) = 1$.

E. Summary

To summarize, the proposed HMM similarity measure $S(\lambda\|\lambda')$ can be computed in the following steps:

1) Compute stationary distributions π and π' for λ and λ'.
2) Compute $D(\mathbf{b}_i\|\mathbf{b}'_{i'})$ for every pair of states $v_i \in \lambda$ and $v'_{i'} \in \lambda'$, using one of the dissimilarity measures given in Section II-C.
3) Evaluate $S_e(v_i, v'_{i'})$ as defined in (6).
4) Estimate the state correspondence matrix $\mathbf{Q}$ as in (4).
5) Compute the HMM similarity measure $S(\lambda\|\lambda')$ as defined in (7), based on the the sparsity measure in (9).

III. Experimental Results

In this section, we present several numerical experiments demonstrate the effectiveness of the proposed similarity measure. We report the results for six different combinations of $S_e(v_i, v'_{i'})$ and $D(.\|.)$, where we consider both $S_e(v_i, v'_{i'}) = 1/D(\mathbf{b}_i\|\mathbf{b}'_{i'})$ and $S_e(v_i, v'_{i'}) = e^{-kD(\mathbf{b}_i\|\mathbf{b}'_{i'})}$ and choose $D(.\|.)$ among the symmetric KLD, Jensen-Shannon divergence (JS), and Bhattacharyya distance (BTC). To compare the performance of our measure with the KLD between two HMMs, in each experiment, we also report symmetric KLD for an observation sequences of length $T = 10,000$.

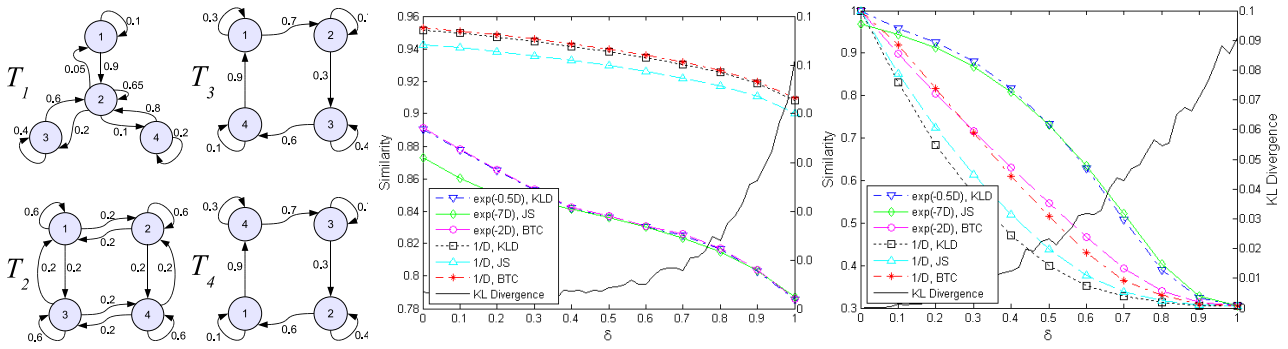
A. Effect of Changing the Model Parameters

First, we study the effect of varying the state transition matrix. Consider the two Markov chains shown in Fig.1(a) with transition matrices $\mathbf{T_1}$ and $\mathbf{T_2}$. The matrix $\mathbf{T_1}$ represents a Markov chain, in which the second state has a higher probability of being visited compared to the other states, while $\mathbf{T_2}$ represents a Markov chain with a uniform stationary distribution. We define the transition probability matrices of λ and λ' as $\mathbf{A} = \mathbf{T_1}$ and $\mathbf{A}' = (1-\delta)\mathbf{T_1} + \delta\mathbf{T_2}$, respectively. So $\mathbf{A}$ is fixed, while $\mathbf{A}'$ changes from $\mathbf{T_1}$ to $\mathbf{T_2}$ by varying δ from 0 to 1. We define the emission probability matrices $\mathbf{B} = [b_{i,j}]$ and $\mathbf{B}' = [b'_{i,j}]$ as:

$$\mathbf{B} = \begin{bmatrix} 0.25 & 0.75 & 0 & 0 \\ 0.35 & 0.65 & 0 & 0 \\ 0 & 0 & 0.15 & 0.85 \\ 0 & 0 & 0.45 & 0.55 \end{bmatrix}, \mathbf{B}' = \begin{bmatrix} 0.2 & 0.8 & 0 & 0 \\ 0.3 & 0.7 & 0 & 0 \\ 0 & 0 & 0.1 & 0.9 \\ 0 & 0 & 0.4 & 0.6 \end{bmatrix},$$

where $b_{i,j}$ and $b'_{i,j}$ are the probability of emitting the jth symbol at the ith state in the respective HMMs. As we see in Fig.1(b), the similarity between the two HMMs decreases as δ increases, and as a result, as $\mathbf{A}'$ diverges from $\mathbf{A}$.

Next, we study the effect of changing the emission probability matrix. In this experiment, we let $\mathbf{A} = \mathbf{A}' = \mathbf{T_3}$ (defined in Fig.1(a)) and define $\mathbf{B} = \mathbf{E_1}$ and $\mathbf{B}' = (1-\delta)\mathbf{E_1} + \delta\mathbf{E_2}$, where $\mathbf{E_1} = \mathbf{I}_{4\times4}$ and $\mathbf{E_2} = 0.25 \cdot \mathbf{1}_{4\times4}$. Here, $\mathbf{I}$ denotes the identity matrix and $\mathbf{1}$ denotes an all-one matrix. Therefore, $\mathbf{B}'$ for λ' changes from deterministic symbol emission ($\mathbf{B}' = \mathbf{E_1}$) to a uniformly distributed emission ($\mathbf{B}' = \mathbf{E_2}$). We can see a similar trend in Fig.1(c) as before, where the similarity between λ and λ' effectively decreases for an increasing δ.



(a) (b) (c)

Fig. 1. Effect of changing the HMM parameters on the proposed similarity measure: (a) Markov chains used in the experiments. (b) Changes in the similarity measure when varying the transition probability matrix. (c) Similarity changes when varying the emission probability matrix.

To see how similarity affects the sparsity of the state correspondence matrix, let us consider two points $\delta = 0$ and $\delta = 1$ in Fig. 1(c) (for "$\exp(-0.5D)$,KLD"):

$$\mathbf{Q}_0 = \begin{bmatrix} 0.12 & 0 & 0 & 0 \\ 0 & 0.65 & 0 & 0 \\ 0 & 0 & 0.16 & 0 \\ 0 & 0 & 0 & 0.07 \end{bmatrix}, \mathbf{Q}_1 = \begin{bmatrix} 0.04 & 0.08 & 0.04 & 0.03 \\ 0.08 & 0.19 & 0.10 & 0.07 \\ 0.04 & 0.10 & 0.05 & 0.03 \\ 0.03 & 0.07 & 0.03 & 0.02 \end{bmatrix}$$

As expected, $\mathbf{Q_0}$, corresponding to similar HMMs ($\delta = 0$), is a sparse matrix, while $\mathbf{Q_1}$, for distant matrices ($\delta = 1$), is not.

B. Effect of Permuting the States

In this experiment, we define $\mathbf{A} = \mathbf{T_3}$, $\mathbf{A}' = (1-\delta)\mathbf{T_3} + \delta\mathbf{T_4}$, $\mathbf{B} = \mathbf{E_1}$, and $\mathbf{B}' = (1-\delta)\mathbf{E_1} + \delta\mathbf{E_2}$, where $\mathbf{T_3}$ and $\mathbf{T_4}$ are the transition matrices of two right Markov chains shown in Fig. 1(a), and $\mathbf{E_1} = \mathbf{I}_{4\times4}$ and $\mathbf{E_2} = \mathbf{I'}_{4\times4}$, where $\mathbf{I'}_{4\times4}$ denotes the 90° rotated version of the identity matrix $\mathbf{I}_{4\times4}$.

As we can see, for $\delta = 0$, the two HMMs are identical. When we increase δ, λ' diverges from λ, but when we reach $\delta = 1$, λ' becomes essentially identical to λ except that its states are permuted. Therefore, we expect that the similarity between λ and λ' will be maximized at both $\delta = 0$ and $\delta = 1$. In fact, this can be observed in Fig.2.

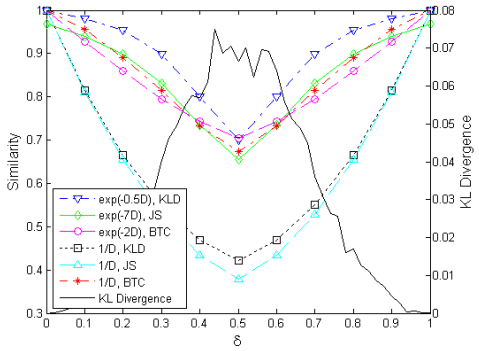


Fig. 2. Effect of permuting the states.

C. Comparing Multivariate Gaussian HMMs

Finally, we demonstrate the effectiveness of the proposed similarity measure for comparing HMMs whose emission probabilities follow a multivariate Gaussian distribution. In this experiment, we use the symmetric KLD as the dissimilarity measure between states in the respective HMMs, utilizing the closed-form expression of KLD for two d-dimensional Gaussian distributions as given in [2].

Here, we let $\mathbf{A} = \mathbf{A}' = \left(\begin{smallmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{smallmatrix}\right)$ and define the emission probability distributions of the two HMMs as multivariate Gaussian with parameters $\boldsymbol{\mu}_1 = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, $\boldsymbol{\mu}_2 = 3\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_1' = \left(\begin{smallmatrix} 1 \\ 3-\delta \end{smallmatrix}\right)$, $\boldsymbol{\mu}_2' = \left(\begin{smallmatrix} 3 \\ 1+\delta \end{smallmatrix}\right)$, $\mathbf{C}_1 = \mathbf{C}_1' = \mathbf{I}_{2\times 2}$, $\mathbf{C}_2 = \left(\begin{smallmatrix} 1 & 0.3 \\ 0.3 & 1 \end{smallmatrix}\right)$, and $\mathbf{C}_2' = \left(\begin{smallmatrix} 1 & 0.1 \\ 0.1 & 1 \end{smallmatrix}\right)$. The effect of changing δ from 0 to 2 is shown in Fig.3. As we expect, we can see that the similarity increases as δ increases.

D. Comparison with KLD

As we can see in the previous simulation results, even the use of long observation sequences with $T = 10,000$ symbols are generally not enough to get an accurate estimate of KLD. Unlike KLD, the proposed similarity measure can be accurately estimated without any random fluctuations, and it can be computed around a thousand times faster in Matlab.

IV. Conclusion

In this work, we proposed a novel low-complexity HMM similarity measure. The proposed scheme first estimates the correspondence score between states in different HMMs, based on the concept of semi-Markov random walk. The similarity between the given HMMs can then be measured by evaluating the sparsity of the state correspondence score matrix. Estimation of the proposed measure does not require Monte-Carlo simulations, hence can be accurately estimated in a much more efficient manner, compared to the conventional KLD. The low-complexity of

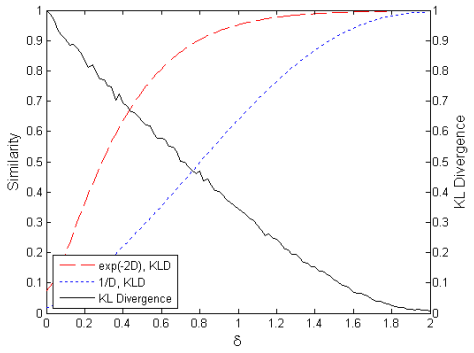


Fig. 3. Comparing multivariate Gaussian HMMs.

the proposed approach makes this novel similarity measure especially suitable for applications with large datasets or complex HMMs, or those that require real-time computing. The proposed measure is bounded, symmetric, and can be effectively implemented for both HMMs with discrete or continuous emission probabilities.

REFERENCES

[1] B.-H. Juang and L. R. Rabiner, "A probabilistic distance measure for hidden markov models," *AT&T Technical Journal*, vol. 64, no. 2, pp. 391 – 408, 1985.

[2] M. Do, "Fast approximation of kullback-leibler distance for dependence trees and hidden markov models," *Signal Processing Letters, IEEE*, vol. 10, no. 4, pp. 115 – 118, apr 2003.

[3] M. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden markov models," *Multimedia, IEEE Transactions on*, vol. 4, no. 4, pp. 517 – 527, dec 2002.

[4] L. Chen and H. Man, "Fast schemes for computing similarities between gaussian hmms and their applications in texture image classification," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 1984–1993, 2005.

[5] C. Bahlmann and H. Burkhardt, "Measuring hmm similarity with the bayes probability of error and its application to online handwriting recognition," in *In Proc. of the 6th ICDAR*, 2001, pp. 406–411.

[6] J. Soding, "Protein homology detection by HMM-HMM comparison," *Bioinformatics*, vol. 21, no. 7, pp. 951–960, 2005.

[7] J. Silva and S. Narayanan, "Upper bound kullback-leibler divergence for transient hidden markov models," *Signal Processing, IEEE Transactions on*, vol. 56, no. 9, pp. 4176 –4188, sept. 2008.

[8] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1958.

[9] R. B. Lyngsø, C. N. S. Pedersen, and H. Nielsen, "Metrics and similarity measures for hidden markov models," in *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, 1999, pp. 178–186.

[10] J. Zeng, J. Duan, and C. Wu, "A new distance measure for hidden markov models," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1550 – 1555, 2010.

[11] S. M. Ross, *Stochastic Processes*, 2nd ed. John Wiley & Sons, Inc., 1996.

[12] S. V. N. Vishwanathan, N. N. Schraudolf, R. Kondor, and K. M. Borgwardt, "Graph kernels," *Journal of Machine Learning Research*, vol. 11, pp. 1201–1242, Apr 2010.

[13] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal Processing*, vol. 18, no. 4, pp. 349 – 369, 1989.

[14] L. Lee, "Measures of distributional similarity," in *37th Annual Meeting of the Association for Computational Linguistics*, 1999, pp. 25–32.

[15] N. Hurley and S. Rickard, "Comparing measures of sparsity," *Information Theory, IEEE Transactions on*, vol. 55, no. 10, pp. 4723 –4741, oct. 2009.