Embedding HMM's-based models in a Euclidean space: The topological hidden Markov models

Bouchaffra, D. ;

This paper appears in: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on
Issue Date : 8-11 Dec. 2008
On page(s): 1 - 4
ISSN : 1051-4651
Print ISBN: 978-1-4244-2174-9
References Cited: 9
INSPEC Accession Number: 10444332
Digital Object Identifier : 10.1109/ICPR.2008.4761135
Date of Current Version : 23 January 2009

ABSTRACT

One of the major limitations of HMM-based models is the inability to cope with topology: when applied to a visible observation (VO) sequence, HMM-based techniques have difficulty predicting the n-dimensional shape formed by the symbols of the VO sequence. To fulfill this need, we propose a novel paradigm named "topological hidden Markov models" (THMM's) that classifies VO sequences by embedding the nodes of an HMM state transition graph in a Euclidean space. We have applied the concept of THMM's to: (i) predict the ASCII class assigned to a handwritten numeral, and (ii) map a protein primary structure to its 3D fold. The results show that the concept of second level THMM's outperforms the SHMM's and the SVM classifiers.

Embedding HMM's-based Models in a Euclidean Space: The Topological Hidden Markov Models

Djamel Bouchaffra
Grambling State University
dbouchaffra@ieee.org

Abstract

One of the major limitations of HMM-based models is the inability to cope with topology: When applied to a visible observation (VO) sequence, HMM-based techniques have difficulty predicting the n-dimensional shape formed by the symbols of the VO sequence. To fulfill this need, we propose a novel paradigm named "topological hidden Markov models" (THMM's) that classifies VO sequences by embedding the nodes of an HMM state transition graph in a Euclidean space. We have applied the concept of THMM's to: (i) predict the ASCII class assigned to a handwritten numeral, and (ii) map a protein primary structure to its 3D fold. The results show that the concept of second level THMM's outperforms the SHMM's and the SVM classifiers.

1. Introduction

The real milestone of the hidden Markov models (HMM's) occurred when applied to speech recognition in the late 1980's [8]. Signal processing [4], and document analysis [9] have also exploited the HMM´s resources. Half a decade later, HMM's spread to many other areas such as image processing, computer vision [7], and biosciences [1].

However, the use of HMM's remains scarce. The main reason behind this limitation is explained by the fact that HMM's are unable to: (i) account for long range dependencies which unfold structural information, and (ii) capture topological features [6] such as the shape formed by the visible observation (VO) sequence. Because the traditional HMM's modeling is based on the hidden state conditional independence assumption of the VO sequence, therefore, HMM's make no use of structure. Furthermore, the fact that the HMM's state transition graph is not embedded in a Euclidean space, therefore HMM's make no use of topology. *This lack of structure and topology inherent to HMM's has drastically limited object recognition.* To overcome this problem, a few numbers of approaches have been proposed. The hierarchical HMM's (HHMM's) introduced in [5] are capable to model complex multi-scale structure which appears in many natural sequences. The structural HMM's (SHMM's) introduced in [3] offer a methodology that automatically identifies the different constituents called "local structures". Nevertheless, this generalization of HMM's to capture local structures *did not address the shape modeling problem of the VO sequence*. The embedding of topological features assigned to local structures within HMM's has rarely been a focus in the machine learning community.

We propose *a novel machine learning paradigm that embeds the nodes of an HMM state transition graph in a Euclidean space*. This new approach entitled topological hidden Markov models (THMM's) extends the traditional HMM's by: (i) modeling the local structures of the entire VO sequence and (ii) extracting their shapes. There are many applications where THMM's can be applied. A first one would be in speech recognition where the pitch contour of some speech units (phonemes, syllables) groupings can be extracted to provide complementary information about the uttered phrase. The fusion of a locale and a global analysis of the signal will enhance speech recognition. A second application would be to classify celestial objects based on morphological features. It is well known that the ages of galaxies are explained in part by the shape formed by their constituents (aggregates of stars, gas and dust). Galaxy classification will leapfrog our understanding about the origin of the universe. Section 2 clarifies the notion of VO sequence. Section 3 depicts the topological mapping between the VO sequence and the shape it forms. Section 4 introduces the THMM's concept. Two applications are presented in Section 5. The conclusion is laid in Section 6.

2. The Visible Observation Sequence

We define a visible observation (VO) sequence as an incoming flow of symbols without a visible structure. However, we define a unit of information (UNIF) as a shape formed by a group of symbols of a VO sequence. For example, a cloud of points representing a VO sequence forms a circle viewed as UNIF. Not all VO subsequences constitute a UNIF;

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

only those disclosing structural constituents that exhibit this pattern. We introduce some applications that are intended to clarify the notions of VO and UNIF sequence. A first one consists of classifying the structure of minerals based on the topology of the bonds that link the atoms in the crystal. For example, the butane gas linear formula "CHHHCHHCHHC-HHH" represents a VO sequence. However, the same formula can be written in a more informative way as a sequence of UNIF's: "$CH_3CH_2CH_2CH_3$". *A UNIF in the butane gas molecule is the shape associated to either the subsequence "CH_3" or "CH_2"* (see figure 1). A second application aims to map handwritten word sequences onto their ASCII representations. A handwritten word sequence (or script) such as "**The quick brown fox**" is viewed as a sequence of pixels. Each isolated character can be categorized as one of the 5 classes "Ascender" (A), "Descender" (D), "Median" (M), "Both Ascender-Descender" (B), and "Space" (S). Since the first handwritten character of this script corresponding to the letter "T" is moving upward, therefore it is depicted as "A". The second handwritten character assigned to the letter "h" is also perceived as "A", whereas the third character assigned to "e" is depicted as "M" since it remains in the median line of the handwritten script. One can finally represent the script "**The quick brown fox**" as the VO sequence "AAMSDMMMASAM-MMMSAMM". However, *it is worth to underscore that a particular "deformation" of a group of symbols (composed of A, M, D) produces a handwritten word with a shape*. Because a word has a potential to convey a meaning, it represents a UNIF. A shape of a handwritten word can be extracted using the "pixel histogram" method (horizontal/vertical scan).

3. Topological Mapping: Projection Onto a Euclidean Space

The thrust is to determine a mapping between a VO sequence and the contour of its UNIF. We assume that the VO sequence selected possesses a "meaningful" structure. To do so, we first map through a function f the VO sequence to its UNIF: this mapping is called a "Sequence Deformation" since the VO segment has been deformed to form a UNIF. We then, map through a function g the UNIF to its shape using a contour representation technique. A *Fourier* or a *Wavelet* coefficient vector $[a_0,a_1,\ldots,a_j]^T$ describing the external contour is computed in this phase. This mapping is called a "Shape Representation". The composite function (g°f) relates the VO sequence $O = o_1,o_2,\ldots,o_T$ to its shape vector defined in an Euclidean space. This mapping allows the traditional HMM to be ingrained in a Euclidean space.

4. Topological Hidden Markov Models

The goal of the THMM's is to map a VO sequence into one class of a finite set of classes. To achieve this goal, we first need to extract the UNIF's from the VO sequence and capture their external contours.

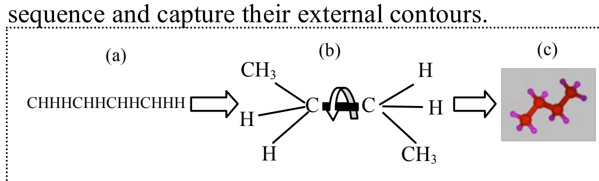

Figure 1. Butane molecule: (a) VO sequence, (b) UNIF sequence, and (c) UNIF shapes.

4.1. UNIF Shape Representation

Shapes of UNIF's are captured by their external contours. A contour is viewed as a discrete signal that consists of low-frequency and high-frequency contents. The low-frequency content is the most important part of the signal, since it provides the signal with its identity: This part is known as the pure signal. However, the high-frequency signal conveys flavor or nuance: This part is associated with noise. The thrust behind the concept of THMM's is to express the probability distribution assigned to the pure signal as a function of the Gaussian distribution assigned to the signal noise. Let $O = o_1,o_2,\ldots,o_T$ be a VO sequence of length T made of symbols o_i. Let $X(t)= \{x(t)\}$ $(t=1,\ldots,m)$ be the closed contour representation of length m that captures the shape of its UNIF. Each point of this contour is designated by $x(t) = [x_1(t), x_2(t),\ldots,x_n(t)]^T$. Our goal is to extract the noisy part of a signal during the shape analysis of the object. Whatever shape representation technique might be adopted, one can coarsely approximate the original signal x(t) by decomposing it into a sum of a pure signal $\zeta(t)$ and a noisy signal N(t): $x(t) = \zeta(t) \oplus N(t)$.

4.2. First Level THMM's

Given a model λ, the VO sequence O, its UNIF contour sequence $X(t)= \{x(t)\}$ $(t=1,\ldots,m)$; evaluate the match between λ, and this VO sequence O by computing $P(O|\lambda)$. If q stands for the hidden state sequence assigned to O, then:

$P(O|\lambda) =\Sigma_q P[O,X(t),q|\lambda]$. Since each noise point on a contour depends only on its symbol o_k, therefore:

$$P(O|\lambda) \approx \sum_{q} [\prod_{t=1}^{t=m} P[N(t) = x(t) - \zeta(t)|o_{x(t)}] \times P(o_t|q_t) \times P(q_t|q_{t-1})],$$

where: $o_{x(t)}$ are the k symbols of the VO sequence such that $f(o) \supset x(t)$, and N(t) is a multivariate Gaussian distribution assigned to the contour noise.

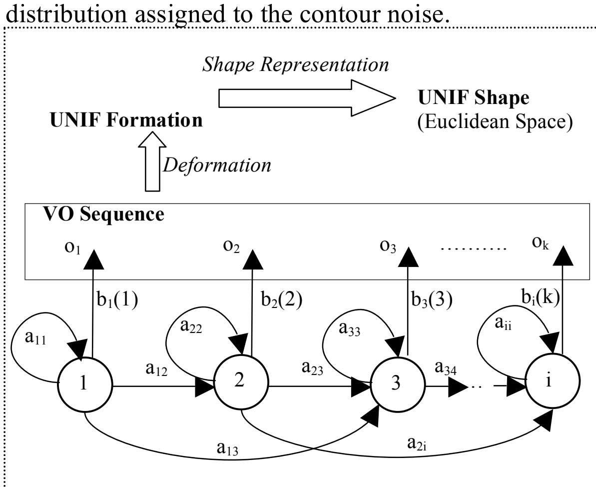

Figure 2. The state transition graph of a first level topological HMM.

Definition 4.1. *A first level THMM is a quadruple* $\lambda = [\pi,A,B,T]$, *where:* $\pi = \{\pi_i\} = P(q_0 = e_i)$ *is the initial hidden state probability vector,* $A = \{a_{ij}\} = P(q_{t+1} = j|q_t = i)$ *is the hidden state transition probability matrix,* $B = \{b_j(k)\} = P(o_k$ *at time* $t|q_t = e_j)$ *is the emission probability matrix,* $T = P[N(t) = x(t) - \zeta(t) \mid o_{x(t)}]$ *is the Gaussian distribution of the noise produced by the k contour points x(t). Figure 2 depicts its state transition graph.*

4.2.1. Problems Assigned to a First Level THMM. **(i) Probability Evaluation:** Given a model λ and a VO sequence O with its corresponding UNIF external contour point sequence $X(t)= \{x(t)\}$ $(t=1,\ldots,m)$, the goal is to evaluate how well does λ match O.
(ii) Statistical Decoding: Find the best hidden state sequence that best "explains" the VO sequence.
(iii) Topological Decoding: Determine the "correct" shape of the UNIF assigned to the VO sequence O via the noise evaluation of its external contour.
(iv) Learning: Estimate the model parameters $\lambda = [\pi,A,B,T]$ that maximize $P(O|\lambda)$.

4.3. Second Level Topological HMM's

Psychophysical studies [2] have shown that humans can recognize objects using fragments of outline contour alone. In this context, a VO sequence $O = o_1, o_2,\ldots,o_T$ is viewed as constituents $O_1,O_2,\ldots,O_s$. Each O_i is a string of symbols $o_i \in \Sigma$ interrelated in some way.

The tasks within the second level THMM's are: (i) segment an entire VO sequence into s "meaningful" pieces, (ii) determine the shape of each UNIF assigned to a segment O_i by embedding it in a Euclidean space, and (iii) compute the probability of the entire VO sequence for a classification purpose.

4.3.1. UNIF Formation. The objective is to unravel the formation of the UNIF entity. *UNIF's are built through a relation of equivalence defined on a set of vectors S representing shapes*. Each class of equivalence is a UNIF that contains shapes whose external contours are similar with respect to some metric distance. Each UNIF U_i describes piecewise the global shape formed by the entire VO sequence. Through this partitioning operated on contours, discrimination of shapes will be achieved.

4.3.2. Formulation of the Second Level THMM's. Let $U = U_1,\ldots,U_s$ be the UNIF sequence assigned to the VO sequence $O = O_1,\ldots,O_s$, and $X(t) = \{X_i(t)\}$ $(i=1,\ldots,s)$ be the sequence of all contours assigned to the UNIF sequence. The probability of the VO sequence O with its contour X(t) given a model λ is: $P(O \mid \lambda) = \sum_U P[O,X(t),U \mid \lambda]$. However, if: $P[X_i(t) \mid O_i] = \prod_{t=1,\ldots,ri} P[N_i(t) \mid O_i] = \phi_i$ and if $\frac{P(U_i|O_i) \times P(O_i) \times P(U_i|U_{i-1})}{P(U_i)} = \Psi_i$, then:

$$P(O|\lambda) \approx \sum_{U_1U_2\ldots U_s} \prod_{i=1}^{i=s} [\phi_i] \times [\Psi_i].$$

Definition 4.2. *A second level THMM extends the first level THMM by incorporating the posterior probability matrix U of a UNIF* U_i *given its constituent* O_i *and the UNIF transition matrix D. It is therefore a sextuple* $\lambda = [\pi,A,B,U,D,T]$.

4.3.3. Problems Assigned to a Second Level THMM. We add to the first level THMM problems the **Structural Decoding** which determines the optimal UNIF sequence $U^* =< U_1^*,U_2^*,\ldots,U_s^* >$ such that: $U^* = \operatorname{argmax}_U P(O,U \mid \lambda)$. For example, a sequence such as: <round, straight,...,convex> is derived to describe the global shape formed by the VO sequence. However, the **Topological Decoding** Computes the "correct" shapes of the UNIF's assigned to the VO sequences O_i's. The sequence <round, straight,...,convex> is decoded in terms of its contour vector sequence.

5. Selected Applications

To demonstrate the significance of the THMM's, we have selected two applications: (i) handwritten numeral recognition, and (ii) protein fold recognition.

5.1. Handwritten Numeral Recognition

Objective: Map a handwritten numeral to one of the 10 ASCII digits. A **VO sequence** is a sequence of 2D coordinate points of the external contour of the numeral. A **UNIF** is a stroke (segment) assigned to a VO subsequence. A **UNIF shape** is a sequence from 8-directions chain code symbols. We compared the THMM's approach with the SHMM's classification technique. The training (60,000 digits) and testing (10,000 digits) were conducted on the MNIST database. We have obtained 98.38% (96.81%) recognition accuracy from THMM's and (SHMM's) classifiers with a rejection rate of about 1.24% and (1.98%), respectively.

Fold	SVM	SHMM	Comb	THMM
4	43.7	35.0	50.0	58.3
9	69.8	50.0	77.8	83.9
11	50.0	66.7	66.7	73.3
26	46.8	34.7	61.5	71.0
30	25.0	33.3	33.3	50.0
33	50.0	75.5	50.0	78.1
51	31.2	30.0	48.1	56.2
…	…	…	…	…
Mean	**45.2**	**51.6**	**61.6**	**71.2**

Table 1. Accuracy (%) of THMM, SVM, SHMM, and combination of SHMM/SVM (Comb) on some protein fold classes λ_i's.

Models	Improvement (%)
(Comb-SVM)/SVM	36.2
(Comb-SHMM)/SHMM	19.3
(THMM-SVM)/SVM	**57.4**
(THMM-SHMM)/SHMM	37.9
(THMM-Comb)/Comb	15.5

Table 2. Relative improvements among classifiers.

5.2. Protein Fold Recognition

Objective: Map an amino acid sequence to one of the 27 protein folds. A **VO sequence** is a linear sequence of amino acids. A **UNIF** is a protein secondary structure (e.g., α-Helix, β-Sheet, etc.). A **UNIF shape** is captured using a "Dual-Tree Complex Wavelet" Transform. We extracted 16 UNIF's using a partition in classes of equivalence. We compared the THMM's approach with other classifiers. The training (605 proteins) and testing (385 proteins) were conducted on the SCOP database. The results depicted by Table 1 show the superiority in performance of the THMM's over other classifiers. The relative improvements are illustrated by Table 2.

6. Conclusion

We have presented a novel machine learning paradigm that embeds the nodes of HMM-based models in a Euclidean space. Our approach decomposes the VO sequence into segments in order to unravel their UNIF's. The UNIF's are generated via a partition of their shapes into classes of equivalence. Therefore, the THMM approach is well-suited to: (i) exploit long-range dependencies, and (ii) account for metric information related to the object depicted by the VO sequence. THMM's extend several HMM's based paradigms that are not adequate to provide an insight into the structural world. Results show that the THMM's concept has significantly outperformed both the SVM, and the SHMM's classifiers. We believe that this embedment of topology within the realm of HMM's will open a new area in which dynamic Bayesian networks can exploit more powerful topological features.

7. References

[1] K. Asai, S. Hayamizu, and H. Handa, "Prediction of protein secondary structures by hidden markov models", *Computer Application in the Biosciences (CABIOS)*, 9(2):141–146, 1993.

[2] I. Biederman and G. Ju, "Surface vs. edge-based determinants of visual recognition", *Cogn. Psych.*, 20(1), 1988.

[3] D. Bouchaffra and J. Tan, "Structural hidden markov models using a relation of equivalence: Application to automotive designs", *Data Mining and Knowledge Discovery Journal*, Springer V., 12(1), 2006.

[4] G. Fan and X. Xia, "Improved hidden markov models in the wavelet-domain", *IEEE TSP*, 49(1), 2001.

[5] S. Fine, Y. Singer, and N. Tishby, "The hierarchical hidden markov models: Analysis and applications", *Machine Learning*, 32(1), 1998.

[6] M. Gemignani, *Elementary Topology*, Second edition. Dover Publications, Inc., New York, 1990.

[7] J. Li, A. Najmi, and R. Gray, "Image classification by a two-dimensional hidden markov model", *IEEE TSP*, 482(2):517, Feb 2000.

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

[9] Z. Sun, W. Jiang, J. Sun, Adaptive Online Multistroke Sketch Recognition based on HMM, Springer Verlag, 2006.