Pattern Recognition 43 (2010) 2590–2607

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Embedding HMMs-based models in a Euclidean space: the topological hidden Markov models

Djamel Bouchaffra *

Department of Mathematics and Computer Science, Grambling State University, LA 71245, USA

ARTICLE INFO

Article history:
Received 6 May 2008
Received in revised form
18 January 2010
Accepted 27 January 2010

Keywords:
Structural hidden Markov models
Structural decoding
Topological decoding
Object contour representation
Protein fold recognition
5×2-fold cross validation paired t-test of hypothesis
Chain code representation
Handwritten numeral recognition

ABSTRACT

Current extensions of hidden Markov models such as structural, hierarchical, coupled, and others have the power to classify complex and highly organized patterns. However, one of their major limitations is the inability to cope with *topology*: When applied to a visible observation (VO) sequence, the traditional 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" (THMMs) that classifies VO sequences by embedding the nodes of an HMM state transition graph in a Euclidean space. This is achieved by modeling the noise embedded in the shape generated by the VO sequence. We cover the first and second level topological HMMs. We describe five basic problems that are assigned to a second level topological hidden Markov model: (1) sequence probability evaluation, (2) statistical decoding, (3) structural decoding, (4) topological decoding, and (5) learning. To show the significance of this research, we have applied the concept of THMMs to: (i) predict the ASCII class assigned to a handwritten numeral, and (ii) map protein primary structures to their 3D folds. The results show that the second level THMMs outperform the SHMMs and the multi-class SVM classifiers significantly.

1. Introduction

The concept of hidden Markov models (HMMs) has been introduced in the sixties by Baum and his colleagues [1]. It is a widely used approach that models time series problems from a statistical view. The real milestone of HMMs occurred when they were applied to speech processing and recognition in the late 1980s [2,3]. Related areas such as signal processing [4,5], and handwriting and text recognition [6–8] have also exploited the resources of these stochastic models. Half a decade later, HMMs spread to many other areas such as image processing, computer vision [9], biosciences [10], and control [11]. Promising results have been obtained from the use of HMMs in several applications in the aforementioned areas. However, the number of problems that HMMs can model is insignificant compared to all problems one may encounter. In other words, the use of HMMs by practitioners remains scarce. The main reason behind this limitation is explained by the fact that HMMs are unable to: (i) account for long-range dependencies which unfold structural[1] information and (ii) capture topological features [12] such as the shape[2] formed by the visible observation (VO) sequence. Because the traditional HMMs modeling is based on the hidden state conditional independence assumption of the visible observations, therefore, HMMs make no use of structure. Furthermore, the fact that the HMM state transition graph is not embedded in a Euclidean space, therefore HMMs make no use of topology. *This lack of structure and topology inherent to standard HMMs has drastically limited the shape recognition task of complex objects.*

To overcome the lack of structure inherent to the traditional HMMs, a few number of approaches have been proposed in the literature. The hierarchical hidden Markov models (HHMMs) introduced in [13] are capable to model complex multi-scale structure which appears in many natural sequences. However, the original HHMM's algorithm is rather complicated since it takes $O(T^3)$ time, where T is the length of the sequence, making it impractical for many domains. To decrease the complexity of the HHMM's algorithm, Murphy and Paskin showed that an HHMM is a special kind of dynamic Bayesian network (DBN), and thereby derive a much simpler algorithm whose complexity is $O(T)$ [14]. This connection between HHMMs and DBNs enabled the complexity of the basic HHMM's algorithm to be reduced further.

The structural hidden Markov models (SHMMs) introduced in [15] offer a methodology that automatically identifies the different constituents of a VO sequence. These constituents known as "local structures" are computed via an equivalence relation defined in the space of the VO subsequences. Other

* Tel.: +1 586 744 3184; fax: +1 318 274 6388.
E-mail address: dbouchaffra@ieee.org

[1] From "structure" which is the way in which parts are arranged, or put together to form a whole.

[2] A shape is any subset $S \subset \Re^n$ with a boundary ∂S, restricted to subsets homeomorphic to a ball.

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.01.022

graphical models such as "coupled HMMs" (CHMMs) [16], factorial HMMs (FHMMs) [17], "event-coupled HMMs" (ECHMMs) [18] and "input–output HMMs" (IOHMMs) [19] that illustrate different architectures have also been proposed in the machine learning community to enhance the capabilities of the standard HMMs. Nevertheless, this generalization of the hidden Markov models to capture local structures *did not address the shape modeling problem of the VO sequence. As far as we are aware, the embedding of topological features* (*e.g.*, *shapes*) *of these local structures within HMMs has not been addressed in the literature*.

An other different approach that contributes in building structures is due to Geman's work in vision. He introduced the "compositionality" operation as an ability to construct hierarchical representations of scenes, whereby constituents are viewed in an *infinite variety* of relational compositions. Amongst all possible composition rules that contain syntactical information, statistical criteria such as MDL (minimum description length) and Gibbs distribution have been used to select the optimal interpretation [20]. However, even if this approach unfolds the optimal scene in a tractable manner, it does not reveal the underlying shape of the objects of the scene.

We propose in this paper *a machine learning paradigm that extends the traditional HMMs by embedding the nodes of the state transition graph in a Euclidean space* [21]. This action allows the recognition of objects that exhibit shapes. This new paradigm entitled *topological hidden Markov models* (THMMs) extends the traditional concept of HMMs by: (i) unfolding the constituents of the entire VO sequence and (ii) capturing their shapes. The first level THMMs extracts the global shape formed by the VO sequence. However, the second level THMMs decomposes the entire VO sequence into segments before capturing their local shapes.

There are several applications where THMMs can be applied: A first one would be in speech recognition where the pitch contour (rise and fall of the voice pitch) assigned to some speech units (phonemes, syllables) groupings will be extracted to provide complementary information about the uttered phrase. We believe that the fusion of a locale and a global analysis of the speech signal will be able to enhance the speech recognition task. 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 (large scale aggregates of stars, gas and dust). The galaxy classification task will certainly leapfrog our understanding about the origin of the universe. A third application consists of predicting a protein 3D fold known as tertiary structure given its primary structure (linear sequence of amino acids). Finally, THMMs can be helpful in remote sensing images such as pollution control, crop inventory that involves monitoring and management over a wide agricultural area or seismic wave analysis for earthquake prediction.

The organization of this paper is as follows: Section 2 clarifies the notion of VO sequence through several examples from different applications. Section 3 depicts the topological mapping between the VO sequence and the shape it depicts. Section 4 provides a brief description of the traditional HMMs. The structural hidden Markov model formalism is the object of Section 5. The novel concept of topological hidden Markov models is introduced in Section 6. We cover the first level THMMs, the optimal segmentation problem, and the second level THMMs. Two applications are presented in Section 7. Finally, the conclusion is laid in Section 8.

2. The visible observation sequence

The notion of *visible observation* sequence has been used in many different contexts in the pattern recognition and machine learning community. However, a rigorous definition and the scope of this notion have been often overlooked; they have rarely been addressed thoroughly by researchers. We define a VO sequence as a flow of symbols ordered by time. However, we define a *unit of information* (UNIF) as a shape formed by a group of symbols. If the entire VO sequence has a shape, therefore its shape represents a UNIF that we call *object*. However, if the VO sequence is made of subsequences that possess shapes, therefore each shape is by itself a UNIF. In this case, the sequence of UNIF's obtained represents an entire object. The representation of the UNIF shape is projected into a Euclidean space. A UNIF can unfold only through a *meaningful* organization of the VO sequence. In other words, not all VO sequences constitute a UNIF but only those which disclose structural constituents of the observed object. We introduce some applications from different areas that are intended to clarify the notions of VO and UNIF. A first application would consist 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 "CHHHCHHCHHCHHH" represents a VO sequence; the two symbols "C" and "H" located at different positions span the entire observation sequence. However, the same formula can be written in a more informative way as a sequence of UNIFs: "CH_3–CH_2–CH_2–CH_3". In this formulation, the shapes of the structural parts of the butane which are "CH_3" and "CH_2" are emphasized. A UNIF in the butane gas molecule is the shape associated to either the subsequence "CH_3" or "CH_2". *The UNIF's are certain rearrangements of their constituents that produce shapes* (*refer to* Fig. 1).

A second application is in the area of handwriting recognition: it consists of mapping handwritten word sequences into their ASCII representations. A handwritten word sequence (or script) such as: "The quick brown fox jumps over a sleazy dog" is viewed as a sequence of pixels. However, after several data processing phases including word segmentation, the VO sequence unfolds. Each isolated character can be categorized as one of the five classes "Ascender" (A), "Descender" (D), "Median" (M), "Both Ascender–Descender" (B), and "Space" (S). These classes used in the document analysis area are usually predetermined via an unsupervised clustering algorithm. Since the first handwritten character of this script that corresponds to the letter "T" is rising up (or 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. Following the same procedure, we can finally represent the script: "The quick brown fox jumps over a sleazy dog" as the VO sequence "AAMSDMMMASAMMMMSAMMSDMMDMSMMMM-SAAMSMAMMMDSAMD". This VO sequence transcription is not unique; it is simply one possible manner of globally discerning handwritten phrases. However, *it is worth to underscore that a particular* "*instantiation*" *of a group of symbols made of A, M, D and their connection produces a handwritten word with a shape. Because a word has a potential to convey a meaning, it represents a UNIF* (*refer to* Fig. 2). A shape of a handwritten word can be extracted

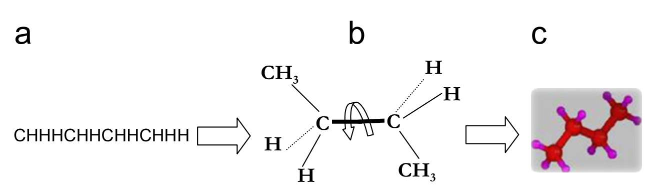

Fig. 1. The butane molecule where: (a) represents its VO sequence O = CHHHCHHCHHCHHH. Each symbol is either a carbon or a hydrogen atom. Part (b) depicts the VO sequence instantiation into UNIF's and (c) outlines UNIF shapes captured by their external contours.