Pattern Recognition 41 (2008) 3092 – 3103

Contents lists available at ScienceDirect

Pattern Recognition

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

Recognition of degraded characters using dynamic Bayesian networks

Laurence Likforman-Sulem*, Marc Sigelle

TELECOM Paris Tech/TSI and CNRS LTCI UMR 5141, 46 rue Barrault F-75634 Paris Cedex 13, France

ARTICLE INFO

Article history:
Received 1 March 2007
Received in revised form 15 January 2008
Accepted 15 March 2008

Keywords:
Markovian models
Hidden Markov models
Dynamic Bayesian networks
Historical documents
Broken character recognition

ABSTRACT

In this paper, we investigate the application of dynamic Bayesian networks (DBNs) to the recognition of degraded characters. DBNs are an extension of one-dimensional hidden Markov models (HMMs) which can handle several observation and state sequences. In our study, characters are represented by the coupling of two HMM architectures into a single DBN model. The interacting HMMs are a *vertical* HMM and a *horizontal* HMM whose observable outputs are the image columns and image rows, respectively. Various couplings are proposed where interactions are achieved through the causal influence between state variables. We compare non-coupled and coupled models on two tasks: the recognition of artificially degraded handwritten digits and the recognition of real degraded old printed characters. Our models show that coupled architectures perform more accurately on degraded characters than basic HMMs, the linear combination of independent HMM scores, as well as discriminative methods such as support vector machines (SVMs).

1. Introduction

Since the seminal work of Rabiner [1], stochastic approaches such as hidden Markov models (HMMs) have been widely applied to speech recognition, handwriting [2,3] and degraded text recognition [4,5]. This is largely due to their ability to cope with incomplete information and non-linear distorsions. These models can handle variable length observation sequences and offer joint segmentation and recognition which are useful to avoid segmenting cursive words into characters [6]. However, HMMs may also be used as classifiers for single characters [7,8] or characters segmented from words by an "*explicit*" segmentation method [9]: the scores output for each character and each class are combined at the word level. Another property of HMMs is that they belong to the class of generative models. Generative models better cope with degradation since they rely on scores output for each character and each class while discriminative models, like neural networks and support vector machines (SVMs), are powerful to discriminate classes through frontiers. In case of degradation, characters are expected to be still correctly classified by generative models even if lower scores are given.

Noisy and degraded text recognition is still a challenging task for a classifier [10]. In the field of historical document analysis, old printed documents have a high occurence of degraded characters, especially broken characters due to ink fading. When dealing with broken characters, several options are generally considered: restoring and enhancing characters [11–13] or recovering characters through sub-graphs within a global word graph optimization scheme [14]. Another solution is to combine classifiers or to combine data. Several methods can be used for combining classifiers [15], one of them consists of multiplying or summing the output scores of each classifier. In the works of [16,17], two HMMs are combined to recognize words. A first HMM, modeling pixel columns, proposes word hypotheses and the corresponding word segmentation into characters. The hypothesized characters or sub segments are then given to a second HMM modeling pixel rows. This second HMM normalizes and classifies single characters. The results of both HMMs are combined by a weighted voting approach or by multiplying scores. Our approach differs with restoration methods as it aims at enhancing the classification of characters without restoration. This is motivated by the fact that preprocessing may introduce distortions to character images. In our previous work [18], we compared data and decision fusion and showed that data fusion yields better accuracy than decision fusion for HMM-based printed character recognition. The present dynamic Bayesian network (DBN) approach is a data fusion scheme which couples two data streams, image columns and image rows into a single DBN classifier. It differs from the approach presented in [16,17] where two classifiers are coupled (one classifier per stream) in a decision fusion scheme, and from a data fusion scheme consisting of a multi-stream HMM which would require large and full covariance matrices in order to take into account dependencies between the streams [18].

Our study consists of building DBN models which include in a single classifier two sequences of observations: the pixel rows and

* Corresponding author. Tel.: +33 1 45 81 73 28.
E-mail address: laurence.likforman@telecom-paristech.fr (L. Likforman-Sulem).

0031-3203/$30.00 © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.03.022

the pixel columns. It can be seen as coupling two HMMs into a single DBN classifier, as opposed to combining the scores of two basic HMM classifiers in a decision fusion scheme. The two HMM architectures, each including an observation stream associated with state variables, are linked in a graphics-based representation. Two different streams are jointly observed and the model parameters (state transition matrices) reflect the spatial correlations between these observations.

We apply the DBN models to broken character recognition. As generative models, DBNs are adapted to degraded character recognition. These models also provide a certain robustness to degradation due to their ability to cope with missing information. They have the ability to exploit spatial correlations between observations. Thus a corrupted observation in the image can be compensated by an uncorrupted one. We compare several DBN architectures among themselves, with other fusion models like the combination of independent HMMs, and with a SVM classifier.

The paper is organized as follows. In Section 2, we briefly introduce Bayesian networks (BN) and DBNs. In Section 3, we present several independent or coupled models. In Section 4, we apply these models to the problem of broken character recognition (artificial and real). We conduct several experiments to show the advantages of DBNs by comparing their performance with the combination of HMM scores and with a SVM classifier. Conclusions are drawn in Section 5.

2. Dynamic Bayesian networks

A (static) BN associated with a set of random variables $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ is a pair: $B = (G, \theta)$ where G is the structure of the BN i.e., a directed acyclic graph (DAG) whose nodes correspond to the variables $X_i \in \mathbf{X}$ and whose edges represent their conditional dependencies, and θ represents the set of parameters encoding the conditional probabilities of each node variable given its parents. The distributions are represented either by a conditional probability table (CPT) when a node and its parents represent discrete variables, or by a conditional probability distribution (CPD) when a node represents a continuous variable. Each CPD usually follows a Gaussian probability density function (pdf). A key property of BNs is that the joint probability distribution factors as

$$P(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} P(X_i \mid Pa(X_i))$$

where $Pa(X_i)$ denotes the parents of X_i. This property is central in the development of fast inference algorithms. Static BNs have been applied to on-line character recognition and signature authentication for modelling dependencies between stroke positions or signature components [19–21].

DBNs are an extension of static BNs to temporal processes occuring at discrete times $t \geqslant 1$. In the following, we consider DBN models which have two observation streams. We will use indices $i = 1, 2$ to denote the two streams. The variables X_i and Y_i denote the respective hidden state and observation attributes in stream i. X_t^i and Y_t^i are the random variables (nodes) for X_i and Y_i at time t.

We assume that the process modelled by DBNs is first-order *Markovian* and stationary. In practice, this means that the parents of any variable X_t^i or Y_t^i belong to the time-slice t or $t-1$ only, and that model parameters are independent of t. Parameters are thus tied and a DBN can be represented by the first two time slices as in Fig. 1. For each observation sequence, the network is repeated as many times as necessary. Fig. 1 shows an example of unrolled DBN for an observation sequence of length $T = 3$: the initial network is repeated T times. Parameters for this model are given by CPTs and CPDs: the three CPTs are the initial state distribution encoding $P(X_1^1)$, the conditional state distribution $P(X_t^2 \mid X_t^1)$, the state transition distribution $P(X_t^2 \mid X_{t-1}^2)$ and the two CPDs are the Gaussian pdfs $P(Y_t^i \mid X_t^i)$, $i = 1, 2$.

DBNs provide general-purpose training and decoding algorithms based on the expectation-maximization (EM) algorithm and on inference mechanisms [22]. Model training consists of estimating model parameters, CPTs and CPDs. Inference algorithms are performed on the network to compute the best state sequences or the likelihoods of observation sequences.

An HMM is a particular case of DBN where there is only one observation stream and one state sequence. The dynamic character of DBNs makes it suitable for applications such as speech and character recognition. In [23,24], DBNs are used to model the interactions between speech observations at different frequency bands in a way that is robust with respect to noise.

3. Independent and coupled architectures

In this study, we couple data streams into single DBN classifiers.This coupling is performed through various DBN architectures (graphical representations) which combine two basic HMMs: the *vertical* HMM whose outputs are the columns of pixels and the *horizontal* HMM whose outpouts are the image rows. In our models, the interactions are usually (but not only) performed through states, leading to efficient models in terms of model complexity (see Section 3.3). Brand et al. [25] have proposed coupled architectures "coupled HMMs" for modeling human interactions: in their models, a state of one HMM is linked to all other HMM states of the adjacent time-slice. This yields symmetric architectures while our coupled architectures are highly non-symmetric.

In our framework, all character classes share the same DBN architecture. Admissible architectures do not include continuous variables with discrete children (for exact inference purposes [23]) and have also a small number of parameters (in order to get a tractable inference algorithm). One approach consists of learning network architecture from data [26]. This approach is tractable for static BNs when dealing with a few observed variables but becomes rapidly too complex in the presence of hidden state variables. Automatic architecture learning is beyond the scope of this paper and our strategy consists of heuristically looking for various admissible architectures and selecting those which provide the best recognition performance.

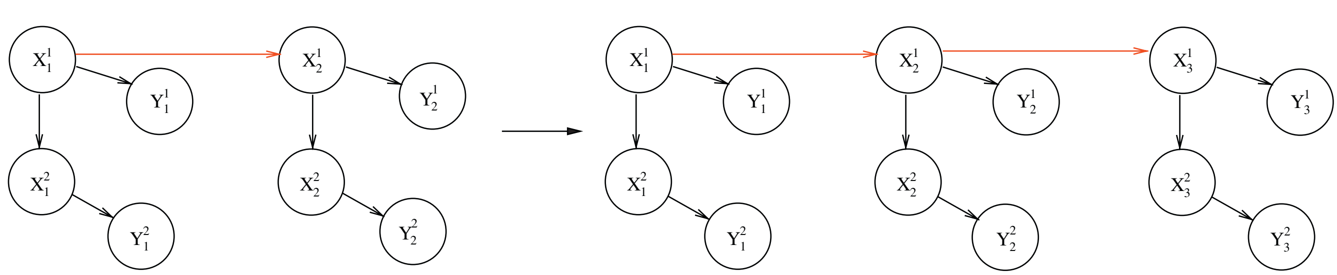

Fig. 1. Because of parameter tying, a DBN can be represented by only two time slices (left). To fit the two observation sequences $\{Y^1\}$ and $\{Y^2\}$ of length $T = 3$, the DBN is unrolled and represented on 3 time slices (right).