Expert Systems With Applications xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Diversified learning for continuous hidden Markov models with application to fault diagnosis

Q1 Zefang Li [a,b], Huajing Fang [a,*], Ming Huang [a]

[a] The School of Automation, Huazhong University of Science and Technology, 1037 Luo-yu Road, Wuhan 430074, PR China
[b] Institute of Computer Technology and Software Engineering, Wuhan Polytechnic, 463 Guan-shan Road, Wuhan 430047, PR China

ARTICLE INFO

Keywords:
Continuous hidden Markov models
Fault diagnosis
Diversified gradient descent algorithm
Likelihood-based model averaging

ABSTRACT

The learning problem of continuous hidden Markov models (CHMMs) is the most critical and challenging one for the application of CHMMs. This paper aims to attack the learning problem of CHMMs by using the diversified gradient descent (DGD) algorithm. The novel learning formula of CHMM parameters, requiring no special form of the objective function and yielding various parameter estimates with different degree of diversity, is derived through dynamically adjusting the iterative procedure according to the gradient change of each parameter. It is the first work for standard CHMM attempting to obtain more local maxima so that the global maximum of the likelihood function of CHMM can be better approximated or even discovered. Hence this paper takes an important step forward in solving the learning problem of CHMM. Furthermore, a likelihood-based model averaging (LBMA) estimator is developed to achieve robust parameter estimation of CHMM based upon the diversiform models attained by the DGD algorithm. The proposed methods are tested on simulation and real-life bearing fault diagnosis problem. The results show that proposed methods perform better in parameter estimation and bearing fault diagnosis compared to the conventional methods.

1. Introduction

In recent years, we have witnessed an increasing interest in continuous hidden Markov models (CHMMs) since they own the powerful capability to model continuous time-varying signals appearing in many real-life applications, such as speech recognition (Chatzis, 2010), market analysis (Dias & Ramos, 2014; Elliott, Siu, & Fung, 2014), computer science (Tosun, 2014), target classification (Robinson, Azimi-Sadjadi, & Salazar, 2005), as well as fault detection and diagnosis (Georgoulas et al., 2013; Geramifard, Xu, Zhou, & Li, 2012; Yuwono, Qin, Zhou, Guo, Celler, & Su, 2015; Zhou, Chen, Dong, Wang, & Yuan, 2015). In the CHMM applications, the learning (or training) problem which still remains open is the most difficult and important issue. The Baum–Welch (BW) algorithm, also known as a special case of the expectation maximization algorithm, is a conventional efficient training algorithm. It adjusts the model parameters according to the Maximum Likelihood (ML) criterion, leading to one local optimum in the parameter space (Rabiner, 1989). Nevertheless, its application is confined to the particular form of the objective function, i.e., the likelihood function (Baum & Egon, 1967). In fact, there are a lot of cases where the requirements of the BW algorithm are evidently not satisfied, such as maximum mutual information training, minimum discrimination information training, and other discriminative training algorithms that have become an active research area recently in speech recognition (Biem, 2006; Huo & Chan, 1993; Liu, Liu, Jiang, Song, & Wang, 2008).

The gradient techniques, which have been shown to obtain solutions comparable to that of the BW algorithm, can also solve for the optimization of the objective function by viewing the learning problem as a classic optimization problem. Unlike the BW algorithm, the gradient techniques can be applied to different objective functions. For example, in Bahl, Brown, de Souza, and Mercer (1986), the gradient descent (GD) method was first used to optimize the maximum mutual information objective function; in (Robinson et al., 2005), a batch gradient descent based method was introduced to estimate the hidden Markov model (HMM) parameters with the ML criterion, and then tested on a multi-aspect under-water target classification problem; in Huo and Chan (1993), a gradient projection method was adopted for HMM training without constraints on the form of the objective function, whose main idea is to search along the projection of the gradient of the objective function on the HMM constraint space for a local maximum. But the efficacy of this method is not validated through experiments. The weakness of the BW algorithm and the gradient techniques mentioned above is that they merely generate the sole searching path in the parameter space on which the ending point corresponds to one single local maximum of the objective

* Corresponding author. Tel.: +86 18971062985.
E-mail addresses: 408729494@qq.com (Z. Li), hjfang@mail.hust.edu.cn (H. Fang), 35706916@qq.com (M. Huang).

http://dx.doi.org/10.1016/j.eswa.2015.08.027
0957-4174/© 2015 Published by Elsevier Ltd.

function. Nevertheless, it is really difficult to find literature focused on the learning problem of standard CHMM during recent 10 years. In this paper the diversified gradient descent (DGD) algorithm for the learning problem of standard CHMM is proposed to obtain diverse parameter estimates and find out more local maxima. To the best of the authors' knowledge, proposed approach is unique in the sense that it is the first idea so far to achieve more local maxima so that the global maximum of the likelihood function of standard CHMM can be better approximated or even discovered.

Conventional HMM-based recognizers focus on building one single model to represent each class, just as if the selected model is known a priori. It is well known that the ML criterion may converge to the true model only when a sufficiently large number of data is available (Nadas, 1983). However, in practice the amount of data is often limited, and a discrepancy exists inherently in the ML-based decoder since the training is conducted on a model-by-model basis while the decoding is carried out by comparing the output of all the models (Biem, 2006). Thereby, the uncertainty exists inevitably in the ML training and decoding process. Furthermore, naive post model selection estimators may underestimate variability, thus lead to overconfident inference and might be unstable. It is usually argued that model averaging can resolve this problem by combining the estimates of numerous good models (Schomaker, 2012; Schomaker & Heumann, 2014; Wang, Zou, & Wan, 2012). The main idea of model averaging is to incorporate the uncertainty associated with the model selection process by means of designing a weighted average across a set of candidate models, hence yields a robust estimation. Until now, in the HMM literature few papers deal with this issue. By making use of diversiform models acquired by the DGD algorithm, this paper proposes a likelihood-based model averaging (LBMA) estimator in order to obtain robust estimation of CHMM parameters.

The rolling bearing is a critical component in almost every rotating and reciprocating machinery where bearing fault distribution varies from about 40% to 90% from large to small machineries depending on the size and type of the machines (Immovilli, Bianchini, Cocconcelli, Bellini, & Rubini, 2013). It is always desired to detect bearing faults as early as possible, and then repair or replace the damaged bearing timely to prevent catastrophic failures and reduce the lengthy industrial downtime. Thereby, much attention has been given to the fault detection and diagnosis of rolling bearings in the past few years (Li, Zhang, & He, 2013; Liang & Faghidi, 2014; Zhao, Jin, Zhao, & Li, 2014). Since the signal generated by a defective bearing is usually non-stationary (Liu, Wang, & Golnaraghi, 2010), the HMM is appropriate for bearing fault diagnosis owing to its great ability to characterize time-varying signals (Zhou et al., 2015). Boutros and Liang (Boutros & Liang, 2011) apply the HMM to diagnose faults successfully in two scenarios: tool wear/fracture and bearing faults; Lebaroud and Clerc (Lebaroud & Clerc, 2008) use the CHMM to classify the stator fault, rotor fault, and bearing fault in induction machines; Yuwono et al. (Yuwono et al., 2015) detect and diagnose bearing defects based on HMM and Swarm Rapid Centroid Estimation which is utilized to estimate the hidden state variables for the HMM.

The fundamental contributions of this paper are as follows. First, it derives the learning formula of CHMM parameters based on the DGD algorithm. The formula has no constraint on the form of the objective function and can acquire diverse transition probabilities as well as observation emission densities, thus yielding more local maxima. Yet the commonly used Baum–Welch algorithm and the gradient descent technique, which can be viewed as a special case of the DGD algorithm, can only obtain one single local maximum of the objective function. Hence, this study provides a better way to approximate or even find out the global maxima of the objective function of standard CHMM. Second, based upon the diversiform parameter estimates, a likelihood-based model averaging estimator is proposed to achieve a robust model rather than the best model by combining various potentially good parameter estimates. Third, to circumvent the computational complexity and guarantee the reliability of CHMM re-estimation procedure, the self-organizing map (SOM), which is an unsupervised learning neural network, is introduced to reduce the dimension of the input vector. This technique can be very useful in the case that the real-life data is high dimensional and some of its components have strong correlation. Lastly, the proposed methods are successfully applied to real-life bearing fault diagnosis problem and satisfied recognition performance is achieved.

The remainder of this paper is organized as follows: In Section 2, in the context of CHMMs the DGD algorithm is formulated both for general objective function optimization and ML criterion. Then, the LBMA estimator is proposed to provide robust parameter estimation in Section 3. Section 4 describes the application of the proposed methods to the fault diagnosis problem of bearings. Details on feature extraction techniques as well as model training and classification procedure are described, which include an introduction to the novel technique based on the SOM. Section 5 shows the experimental results of the simulation examples and bearing fault diagnosis task, along with the comparison of the new systems against other recognition systems using the same database, followed by our conclusions and future work in Section 6.

2. Diversified parameter estimation of CHMM

2.1. DGD method for general objective function optimization

There are usually two essential issues in the HMM-based recognition. One is the determination of an effective objective function $f(\lambda)$ with respect to the HMM parameter set λ. The other is to find an appropriate method to adjust the current model λ so as to optimize the objective function $f(\lambda)$ (Huo & Chan, 1993). In this paper, our main interest is the latter rather than the former. Currently, there is no known way to analytically solve for the latter issue. Alternatively, an iterative procedure ensuring convergence to a local optimum, such as the BW method or the gradient techniques, is the most widely used method (Rabiner, 1989). However, these conventional methods usually update all the parameters, neglecting the important fact that at each step the extent of every parameter approaching its corresponding optimal value is very different from each other because of various magnitudes of their derivatives. It is likely that at certain step part of the parameters have already arrived at their optimal values, i.e., their derivatives have been zero, while the others are still far away from their optima. If the iterative procedure continues to adapt all the parameters including those already being optimal, consequently, more iterative steps are required and a good solution may be overshot. To address the issue, in this study, the diversified learning formulas are derived for the CHMM parameters based on the DGD algorithm. The main idea of the proposed method, stemming from the work of Friedman and Popescu (2003) about the linear regression problem, is to flexibly select some of the CHMM parameters whose derivatives have been very close to zero not to be updated by setting a gradient-adapting factor during each iterative step.

Let us denote an observation sequence as $O = o_1 o_2 \ldots o_T$ with each point $o_t = (o_{t1}, o_{t2}, \ldots, o_{td})$ being a d-dimensional random observation. Let us assume that the emission probability density of the CHMM can be approximated by Gaussian mixture densities, i.e.,

$$b_i(o_t, \Theta_i) = \sum_{m=1}^{M} c_{im} G(o_t, \mu_{im}, \Sigma_{im}) \tag{1}$$

where c_{im} is the mixture coefficient for the mth mixture in the state i, and G is the Gaussian density with the mean vector $\mu_{im} = (\mu_{im1}, \ldots, \mu_{imd})'$ and $d \times d$ covariance matrix $\Sigma_{im} = (\sigma_{imkl})_{k,l=1}^{d}$ for the mth mixture in the state i. Let $\lambda = \{\pi_i, a_{ij}, \Theta_i\}_{i,j=1}^{N}$ denote the CHMM parameter vector, where π_i is the initial state probability,

Please cite this article as: Z. Li et al., Diversified learning for continuous hidden Markov models with application to fault diagnosis, Expert Systems With Applications (2015), http://dx.doi.org/10.1016/j.eswa.2015.08.027