Pattern Recognition Letters 32 (2011) 2009–2014

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Gaussian Process Dynamical Models for hand gesture interpretation in Sign Language

Nuwan Gamage [a], Ye Chow Kuang [a,*], Rini Akmeliawati [b], Serge Demidenko [c]

[a] Monash University Sunway Campus, Jalan Lagoon Selatan, Bandar Sunway, 46150 Selangor D.E., Malaysia
[b] International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia
[c] RMIT International University Saigon South Campus, 702 Nguyen Van Linh Blvd., District 7, HCMC, Viet Nam

ARTICLE INFO

Article history:
Received 6 January 2011
Available online 6 September 2011
Communicated by G. Borgefors

Keywords:
Gesture interpretation
Gaussian Process Dynamical Model
Gaussian Process
Sign Language
Hidden Markov Model

ABSTRACT

Classifying human hand gestures in the context of a Sign Language has been historically dominated by Artificial Neural Networks and Hidden Markov Model with varying degrees of success. The main objective of this paper is to introduce Gaussian Process Dynamical Model as an alternative machine learning method for hand gesture interpretation in Sign Language. In support of this proposition, the paper presents the experimental results for Gaussian Process Dynamical Model against a database of 66 hand gestures from the Malaysian Sign Language. Furthermore, the Gaussian Process Dynamical Model is tested against established Hidden Markov Model for a comparative evaluation. A discussion on why Gaussian Process Dynamical Model is superior over existing methods in Sign Language interpretation task is then presented.

1. Introduction

Machine based automatic *Sign Language* (SL) hand gesture interpretation has long been a popular research topic in *Human Computer Interaction* (HCI). Gesture interpretation accuracy depends on many factors including the adopted learning method, hand gesture features, employed device characteristics, etc. Hardware devices such as single and stereo cameras (Brand and Oliver, 1997), depth-aware cameras (Cappé et al., 2005) and wired glove (Eddy, 1996) were used for gesture data acquiring. Features such as hand movements, hand position, hand pose, finger configuration hand silhouettes are commonly extracted from the input data for gesture interpretation (Freeman and Weissman, 1995). Statistical machine learning methods such as *Artificial Neural Networks* (ANN) (Juang and Rabiner, 1991) and *Hidden Markov Model* (HMM) (Rabiner, 1989; Ong and Ranganath, 2005; Microsoft Corporation, 2010) have been applied to hand gesture learning and interpretation.

The suitability of ANN and HMM in gesture recognition tasks have not been questioned in the past. This paper proposes *Gaussian Process Dynamical Model* (GPDM) as an alternative to HMM and ANN for hand gesture recognition in the context of SL translation. GPDM is more transparent and hence more amenable to human interpretation compared to the black-box approaches such as HMM and ANN. Furthermore, the Gaussian regularisation inherent in GPDM framework enables better generalisation performance in gesture recognition task as will be shown in the below sections. The results are validated on a set of gestures from the *Malaysian Sign Language* (MSL), and the performance is compared against the popular HMM approach.

Section 2 briefly reviews HMM and GPDM machine learning methods. Section 3 explains the proposed methodology and the rationale behind it. Section 4 provides extensive details on the experimental validation of the proposed methodology; the results obtained and related discussion. Section 5 discusses possible future investigations. Finally, Section 6 summarises the findings and contributions of this study.

2. Machine learning

Gesture interpretation for Sign Languages can be regarded as a *supervised learning* problem, as the labels of the training and test samples are known prior to the training process. Two of the supervised learning schemes employed in this study are HMM and GDPM.

2.1. Hidden Markov Model (HMM)

Apart from ANN, HMM is the most popular machine learning method in SL translation. Thus, HMM is used as the benchmark (or controlled) method in this study.

* Corresponding author. Tel.: +60 3 5514 6239; fax: +60 3 5514 6207.
E-mail addresses: nuwan.gamage@gmail.com (N. Gamage), kuang.ye.chow@monash.edu, kuang.ye.chow@eng.monash.edu.my (Y.C. Kuang), rakmelia@iium.edu.my (R. Akmeliawati), serge.demidenko@rmit.edu.vn (S. Demidenko).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.08.015

A formal definition of HMM is given by Ong and Ranganath (2005),

> *"An HMM is a doubly stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observed symbols."*

HMM is an extension of the *Markov process* model. In a regular Markov process, states are inter-connected with state transition probabilities. In HMM, an additional layer of stochastic process (so-called *observations*) is introduced. Observations are derived based on state transition patterns and there are probabilities of occurrence associated with them. Observations are the only visible part in HMM, while the actual Markov process is hidden underneath. As the current state of the Markov process is unobservable at any given moment; it is given the name *'Hidden Markov Model'* (Rabiner and Juang, 1986; Pavlovic et al., 1997; Ong and Ranganath, 2005). Fig. 1 illustrates the operation of HMM graphically.

The challenge in training a HMM is to find a set of suitable model parameters ($\boldsymbol{N},\boldsymbol{M},\boldsymbol{A},\boldsymbol{B},\boldsymbol{\pi}$) that will describe system dynamics satisfactorily while avoiding over-learning. Many variations of basic HMM and advanced algorithms have been developed to achieve this goal. Interested readers can find out the details in (Ong and Ranganath, 2005) and the references there in. Historically, variants of HMM have dominated the gesture interpretation domain. An extensive list of studies in this field is available in (Cappé, 2001; Ong and Ranganath, 2005).

HMMs with mixture of Gaussians outputs (MHMM) reported in (Bilmes, 1998; Murphy, 1998; Fraser, 2008; Resch, 2010) will be used below in the experimental studies. MHMM is one of the latest generalizations of HMM with significant learning capability and flexibility.

2.2. Gaussian Process Dynamical Models (GPDM)

Gaussian Processes (GP) were initially employed for regression of static data (MacKay, 2003). GP alone could not handle tracking, gesture recognition or any other time-series data problems effectively. Furthermore, performing machine learning with high-dimensional data often leads to poor outcomes (Urtasun, 2006). Hence, the *Gaussian Process Latent Variable Model* (GPLVM) was introduced to circumvent these issues (Lawrence, 2003). GPLVM essentially learns the most relevant low-dimensional embedding (or *latent variables*) from the high-dimensional training data (Urtasun, 2006) while discarding statistically less significant variations. The disadvantage of GPLVM is that, there is no description of the relationship between the latent variables. Thus the GPDM was proposed by (Wang and Fleet, 2005, 2008) to address this deficiency of GPLVM.

GPDM can be graphically represented as in Fig. 2(a). The subscript $\boldsymbol{t}$ indicates the time stamp of a time-series data; $\boldsymbol{x_t}$ is the latent variable, $\boldsymbol{y_t}$ is the high dimensional-data; ϕ are basis functions that encode the transition probability from latent variables to each other while $\boldsymbol{\psi}$ are the basis functions that map the latent variables to the training data The mappings are parameterised by $\boldsymbol{A}$ and $\boldsymbol{B}$. These parameters are not of interest from the *Bayesian* perspective of function approximation and may not be unique (Urtasun, 2006).Thus it is possible to marginalize them out as demonstrated in (MacKay, 2003) (this is similar to *Kernel Trick*). Combining all the priors, the latent mapping, and the dynamics provides the final GPDM model as presented in Fig. 2(b). A complete discussion on the GPD mathematical framework is available in (Wang and Fleet, 2005).

GPDM is a fairly recent development in the machine learning space. Its initial applications were mostly concentrated around human motion tracking (Wang and Fleet, 2005; Urtasun, 2006). Hand gesture interpretation of Sign Language is an inherently complex problem, which compounds multidimensional time-series data and fluctuations in different instances (of the same gesture). Moreover, GPDM is a parameter-less model, making it a convenient and a computationally less heavy method in practice. Overall, GPDM brings many advantages in solving certain machine learning problems, mostly in the domain of human motion modelling.

3. Gesture interpretation process

Gesture interpretation process can be divided into three distinct stages, namely: (1) Raw data normalisation; (2) training gesture models; and (3) classification.

3.1. Raw data normalisation

The raw data is captured using a colour digital camera in the video format. A *SwisTrack* (Correll and Sempo, 2006; Lochmatter and Roduit, 2008) implementation of the colour separation and tracking method explained in (Gamage and Akmeliawati, 2009) is used

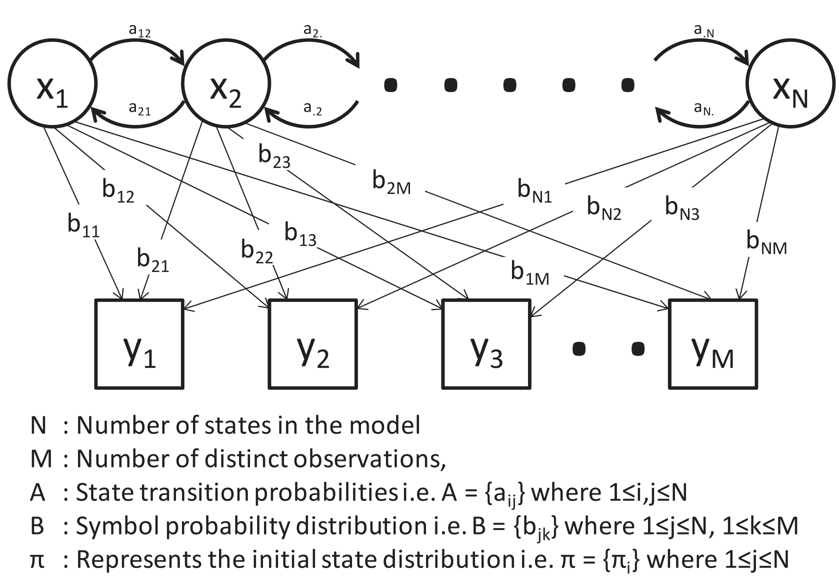

Fig. 1. Hidden Markov Model (x – states, y – observations, a – state transition probabilities, b – output probabilities).