

On Discovering the Correlated Relationship between Static and Dynamic Data in Clinical Gait Analysis

Yin Song¹, Jian Zhang¹, Longbing Cao¹, and Morgan Sangeux²³⁴

¹ Advanced Analytics Institute (AAI), University of Technology, Sydney, Australia

² Hugh Williamson Gait Analysis Laboratory, The Royal Children's Hospital
Melbourne, 50 Flemington Road, Parkville Victoria 3052, Australia

³ Murdoch Childrens Research Institute, Victoria, Australia

⁴ School of Engineering, The University of Melbourne, Victoria, Australia

yin.song@student.uts.edu.au,
{jian.zhang, longbing.cao}@uts.edu.au
morgan.sangeux@rch.org.au

Abstract. ‘Gait’ is a person’s manner of walking. Some patients may have an abnormal gait due to physical impairment. Clinical gait analysis (CGA) is a practical technique and procedure for identifying the underlying impairments that affect the pattern of a patient’s gait. The CGA is critical for further treatment planning. Essentially, CGA tries to use patients’ physical examination results, known as *static* data, to interpret the dynamic characteristics in an abnormal gait, known as *dynamic* data. For many years this process has been qualitatively carried out by gait analysis experts, based mainly on their experience which may lead to subjective diagnoses. To facilitate the automation of this process and form a relatively objective diagnostic view, this paper proposes a new probabilistic correlated static-dynamic model (CSDM) to discover quantitatively correlated relationships between the dynamic characteristics of the gait and their root cause in the static data space. We propose an EM-based algorithm to learn the parameters of the CSDM. One of the main advantages of the CSDM is its ability to provide intuitive knowledge. For example, the CSDM can describe what kinds of static data will lead to what kinds of hidden gait patterns in the form of a decision tree, which helps us to infer what kind of dynamic characteristics would be given to the static data. Our initial experiments indicate that the CSDM is promising for discovering the correlated relationship between the physical examination (static) and gait (dynamic) data.

Keywords: Probabilistic graphical model, Correlated static-dynamic model (CSDM), Clinical gait analysis (CGA), EM algorithm, Decision tree

1 Introduction

The past 20 years have witnessed a burgeoning interest in clinical gait analysis for children with cerebral palsy (CP). The aim of clinical gait analysis is to determine

what is wrong with patients and then plan manageable treatment for them based on the analysis. Usually, two types of data are used in clinical gait analysis: one is the *static* data, which is the physical examination data that is measured when the patient is not walking, such as the shape of the femur and the strength of the thigh muscle. Table 1 shows an excerpt data set from the static data. From the table, we can see that there are many attribute values for each subject. The other type of data is *dynamic* data, which records the dynamic characteristics that evolve during a gait trial and usually can be displayed in curves. Figure 1 shows some example gait curves for one subject. The gait dynamics are recorded from multiple dimensions (i.e., from different parts of the body), such as the pelvis and hips. Since each subject has multiple trials to record the gait, there are multiple curves for each dimension. In addition, each dimension has both the left and right side of the body for recording the dynamic. Thus, the total number of curves for each dimension is the number of trials multiplied by two. We use the red line to denote the dynamic of the left side and the blue line to denote the counterpart of the right side. Figure 1(a)-(d) show 4 different dimensions of the dynamics. Each curve in each dimension represents the corresponding dynamics of one trial for the left or right part. The grey shaded area termed as *normal area* describes the dynamic curve obtained from healthy people with a range of ± 1 standard deviations for each observation point. From the example data shown above, we can see that describing the relationship between the *static* and *dynamic* data in the clinical gait data is not intuitive.

From the perspective of clinical practice, the static data is usually used to explain the existence of abnormal gait in the dynamic data, which is the key task in clinical analysis. In other words, gait analysis experts try to discover the correlated relationships between the *static* and *dynamic* data for further clinic diagnosis. For many years, this process has been conducted empirically by clinical experts and thus is *qualitative*, which may have biased results since the data is complex. Motivated by this, we make an initial exploration in this paper to automatically discover the *quantitative* correlated relationships between the *static* data and *dynamic* curves.

The rest of the paper is organized as following: Section 2 reviews the related work to this paper, followed by the problem formalization in Section 3. Then, Section 4 proposes a probabilistic graphical model to simulate the generating process of the data and gives an EM-based recipe for learning the model given training data. After that, experimental results on both synthetic and real-world data sets are reported in Section 5. Finally, Section 6 concludes this paper.

2 Related Work

In terms of CGA, recent researches [3, 5, 13, 12] have made initial attempts at the automatic discovery of the correlated relationships in clinical gait data by machine learning methods such as multiple linear regression [5] and fuzzy decision trees [12]. However, previous researches have usually preprocessed the gait data and discarded the dynamic characteristics of that data, which fails to explore the

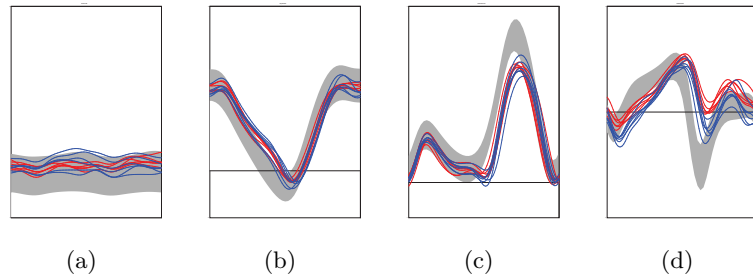


Figure 1: Example Gait Curves for One Patient with 6 Trials: (a) The Pelvic Tilt Dimension; (b) The Hip Flexion Dimension; (c) The Knee Flexion Dimension; (d) The Dorsiflexion Dimension.

Table 1: An Excerpt Data Set from the Static Data

Subject	Internal_Rotation_r	Internal_Rotation_l	Anteversion_r	...	Knee_Flexors_l
1	58	63	25	...	3+
2	60	71	15	...	4
3	53	52	29	...	3
⋮	⋮	⋮	⋮	⋮	⋮

correlated relationship between the static data and the dynamic curves. To the best of our knowledge, our work is the first attempt to explore this correlated relationship comprehensively.

From the perspective of methodologies, there are some existing probabilistic models related to this paper, such as hidden Markov models (HMMs) [11] and conditional random fields (CRFs) [6]. These models focus on modeling the dynamic curves. Thus, they cannot be applied directly here. By contrast, the aim of this paper is to jointly model the static and dynamic data considering their correlated relationships.

3 Problem Statement

Formally, we define the following terms:

- A *static profile* is a collection of static physical examination features of one subject denoted by $\mathbf{y} = (y_1, y_2, \dots, y_L)$, where the subscript i ($1 \leq i \leq L$) denotes the i^{th} attribute of the physical examination features, e.g., the Internal_Rotation_r attribute in Table 1.
- A *gait profile* is a collection of M gait trials made by one subject denoted by $\mathbf{X}_{1:M} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M\}$.

- A *gait trial* (cycle) is multivariate time series denoted by $\mathbf{X}_m = (\mathbf{x}_{m1}, \mathbf{x}_{m2}, \dots, \mathbf{x}_{mN})$, where \mathbf{x}_{mj} ($1 \leq m \leq M$ and $1 \leq j \leq N$) is the j^{th} vector observation of the time series and $\mathbf{x}_{mj} = [x_{m1j} \ x_{m2j} \ \dots \ x_{mDj}]^T$ (D is the number of the dimensions for dynamic data and N is the length of the time series). For example, one dimension of the multivariate time series $(x_{mj1}, x_{mj2}, \dots, x_{mjN})$ ($1 \leq j \leq D$) can be plotted as one curve in Figure 1(a) and represents the dynamics of that dimension for one trial. \mathbf{X}_m can be seen as a collection of such curves in different dimensions.

Our goal is to develop a probabilistic model $p(\mathbf{X}_{1:M}, \mathbf{y})$ that considers the correlated relationships between the *static profile* (i.e., the static data) and the corresponding *gait profile* (i.e., the dynamic data). In other words, we aim to produce a probabilistic model that assigns ‘similar’ data high probability.

4 Proposed Model

4.1 Motivation

The basic idea is to construct the data generating process based on the domain knowledge gained by gait experts and model the process. Specifically, the *static profile* \mathbf{y} of a subject determines the generation of that subject’s potential gait pattern. We denote this hidden gait pattern as a latent variable \mathbf{h} , a vector whose elements h_g ($1 \leq g \leq G$)⁵ are 0 or 1 and sum to 1, where G is the number of hidden gait patterns. The generation of the corresponding *gait profile* $\mathbf{X}_{1:M}$ is then determined by this latent variable \mathbf{h} . In other words, the gait pattern is characterized by a distribution on the gait data. Due to the high dimensionality of $p(\mathbf{X}_{1:M}|\mathbf{h})$, to clarify the generating process of it, we need to consider the corresponding physical process. According to [8], a gait trial can usually be divided into a number of phases and each vector observation \mathbf{x}_{mj} belongs to a certain state indicating its phase stage. These states are usually not labeled and we thus introduce latent variables \mathbf{z}_{mj} ($1 \leq m \leq M, 1 \leq j \leq N_m$) for each vector observation \mathbf{x}_{mj} in each gait trial \mathbf{X}_m . We thus have two advantages: firstly, $p(\mathbf{X}_{1:M}|\mathbf{h})$ can be decomposed into a set of conditional probability distributions (CPDs) whose forms are intuitive to obtain; secondly, the dynamic process of the gait trials are captured by utilizing the domain knowledge.

4.2 The Correlated Static-Dynamic Model

Motivated by the above analysis, we propose a novel correlated static-dynamic model (CSDM), which models the above conjectured data generating process. As mentioned before, existing models (e.g., HMMs and CRFs), cannot be directly used here. This is because HMMs only model the dynamic data $p(\mathbf{X}_m)$ and CRFs only model the relationship between \mathbf{X}_m and \mathbf{z}_m , i.e., $p(\mathbf{z}_m|\mathbf{X}_m)$ ($1 \leq m \leq M$), which is different to our goal of jointly modeling the *static* and

⁵ $h_g = 1$ denotes the g^{th} hidden gait pattern.

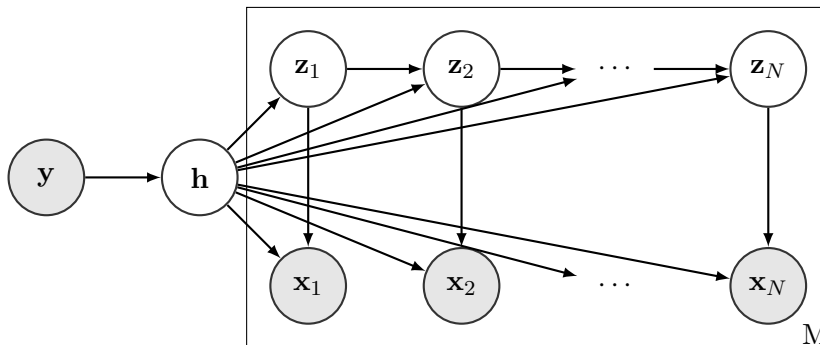


Figure 2: The Graphical Model of the CSDM

gait profiles $p(\mathbf{X}_{1:M}, \mathbf{y})$. The graphical model for the CSDM is shown in Figure 2 (and in Figure 2 we omit the subscript m for convenience). We use conventional notation to represent the graphical model [2]. In Figure 2, each node represents a random variable (or group of random variables). For instance, the *static profile* is represented as a node \mathbf{y} . The directed links express probabilistic causal relationships between these variables. For example, the arrow from the static data \mathbf{y} to the hidden gait pattern variable \mathbf{h} indicates their causal relationships. For multiple variables that are of the same kind, we draw a single representative node and then surround this with a plate, labeled with a number indicating that there are many such kinds of nodes. An example can be found in Figure 2 in which M trials $\mathbf{Z}_{1:M}, \mathbf{X}_{1:M}$ are indicated by a plate label with M . Finally, we denote observed variables by shading the corresponding nodes and the observed static data \mathbf{y} is shown as shaded node in Figure 2. To further illustrate the domain-knowledge-driven data generating process in Figure 2, the generative process for a *static profile* \mathbf{y} to generate a *gait profile* $\mathbf{X}_{1:M}$ is described as follows:

1. Generate the static profile \mathbf{y} by $p(\mathbf{y})$
2. Generate the latent gait pattern \mathbf{h} by $p(\mathbf{h}|\mathbf{y})$
3. For each of the M trials
 - (a) Generate the initial phase state \mathbf{z}_{m1} from $p(\mathbf{z}_{m1}|\mathbf{h})$
 - (b) Generate the corresponding gait observation \mathbf{x}_{m1} by $p(\mathbf{x}_{m1}|\mathbf{z}_{m1}, \mathbf{h})$
 - (c) For each of the gait observations \mathbf{x}_{mn} ($2 \leq n \leq N$)
 - i. Generate the phase state \mathbf{z}_{mn} from $p(\mathbf{z}_{mn}|\mathbf{z}_{m,n-1}, \mathbf{h})$
 - ii. Generate the the corresponding gait observation \mathbf{x}_{mn} from $p(\mathbf{x}_{mn}|\mathbf{z}_{mn}, \mathbf{h})$

4.3 The Parameters of the CSDM

The parameters (i.e., the variables after the semicolon of each CPD) governing the CPDs of the CSDM are listed in the following⁶:

⁶ We assume $p(\mathbf{y}) = \text{const}$ and the const is normalized and determined empirically from the data for convenience. Thus, we do not put it as a parameter.

$$p(\mathbf{h}|\mathbf{y}; \mathbf{d}) = \prod_{g=1}^G d_g(\mathbf{y})^{h_g} \quad (1)$$

where $\mathbf{d}(\cdot)$ is a set of mapping functions with elements $d_g(\cdot) \equiv p(h_g = 1)$ and h_g is the g^{th} element of \mathbf{h} . Since the input \mathbf{y} of the functions is a mixture of discrete and continuous values, it is not intuitive to assume the format of the functions. Thus, here we use the form of a probability estimation tree (PET) [9] to represent the CPD $p(\mathbf{h}|\mathbf{y})$. To be more specific, the parameters governing the CPD is similar to the form “if \mathbf{y} in some value ranges, then the probability of $h_g = 1$ is $d_g(\mathbf{y})$ ”.

$$p(\mathbf{z}_{m1}|h; \boldsymbol{\pi}) = \prod_{g=1}^G \prod_{k=1}^K \pi_{gk}^{h_g, z_{m1k}} \quad (2)$$

where $\boldsymbol{\pi}$ is a matrix of probabilities with elements $\pi_{gk} \equiv p(z_{m1k} = 1|h_g = 1)$.

$$p(\mathbf{z}_{mn}|\mathbf{z}_{m,n-1}, h; \mathbf{A}) = \prod_{g=1}^G \prod_{k=1}^K \prod_{j=1}^K a_{gjk}^{h_g, z_{m,n-1,j}, z_{mnk}} \quad (3)$$

where \mathbf{A} is a matrix of probabilities with elements $a_{gjk} \equiv p(z_{mnk} = 1|z_{m,n-1,j} = 1, h_g = 1)$.

$$p(\mathbf{x}_{ml}|\mathbf{z}_{ml}, h; \boldsymbol{\Phi}) = \prod_{g=1}^G \prod_{k=1}^K p(\mathbf{x}_{ml}|\phi_{gk})^{h_g, z_{mlk}} \quad (4)$$

where $\boldsymbol{\Phi}$ is a matrix with elements ϕ_{gk} . For efficiency, in this paper, we assume that $p(\mathbf{x}_{ml}; \phi_{gk}) = \mathcal{N}(\mathbf{x}_{ml}; \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk})$, which is Gaussian distribution, and thus $\phi_{gk} = (\boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk})$.

Thus, the CSDM can be represented by the parameters $\boldsymbol{\theta} = \{\mathbf{d}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\sigma}\}$.

4.4 Learning the CSDM

In this section we present the algorithm for learning the parameters of the CSDM, given a collection of *gait profiles* $\mathbf{X}_{s,1:M}$ and the corresponding *static profiles* \mathbf{y}_s ($1 \leq s \leq S$) for different subjects. We assume each pair of gait and static profiles are independent of every others since they are from different subjects and share the same set of model parameters. Our goal to find parameters $\boldsymbol{\theta}$ that maximize the log likelihood of the observed data $\mathbf{X}_{1:S,1:M}, \mathbf{y}_{1:S}$ ⁷.

$$L(\boldsymbol{\theta}) = \sum_{s=1}^S \log p(\mathbf{X}_{s,1:M}|\mathbf{y}_s; \boldsymbol{\theta}) \quad (5)$$

⁷ We add the subscript s for representing the s^{th} profile in the rest of the paper.

Algorithm 1: The Learning Algorithm for the Proposed CSDM.

Input : An initial setting for the parameters θ^{old} **Output:** Learned parameters θ^{new}

```

1 while the convergence criterion is not satisfied do
2   |   Estep();
3   |    $\theta^{new} =$  Mstep();
4 end

```

Directly optimizing the above function with respect to θ is very difficult since the involvement of latent variables [2]. Thus, we adopt an expectation-maximization (EM)-based algorithm [4] to learn the parameters, yielding the following iterative method as presented in Algorithm 1. First, the parameters θ^{old} need to be initialized. Then in the E step, $p(\mathbf{z}_{s,1:M}, \mathbf{h}_s | \mathbf{X}_{s,1:M}, \mathbf{y}_s, \theta^{old})$ ($1 \leq s \leq S$) is inferred given the parameters θ^{old} and will be used in M step. The M step then obtains the new parameters θ^{new} that maximize the $Q(\theta, \theta^{old})$ function with respect to θ as follows:

$$Q(\theta, \theta^{old}) = \sum_{s, \mathbf{h}, \mathbf{z}} p(\mathbf{z}_{s,1:M}, \mathbf{h}_s | \mathbf{X}_{s,1:M}, \mathbf{y}_s; \theta^{old}) \log p(\mathbf{h}_s, \mathbf{z}_{s,1:M}, \mathbf{X}_{s,1:M}, \mathbf{y}_s; \theta) \quad (6)$$

The E and M steps iterate until the convergence criterion is satisfied. In this manner, $L(\theta)$ is guaranteed to increase after each interaction.

Challenges of the Learning Algorithms The challenges of the above algorithm then become the calculation of the E step and the M step. A standard forward-backward inference algorithm [11] cannot be directly used here for the E step because of the introduction of latent variables \mathbf{h}_s ($1 \leq s \leq S$). Thus, we provide a modified forward-backward inference algorithm in Algorithm 2 considering the involvement of \mathbf{h}_s ($1 \leq s \leq S$). In calculating the M step, it is difficult to find an analytic solution for $\mathbf{d}(\cdot)$. Thus, we utilize a heuristic algorithm to solve it in Procedure estimatePET. The details of the implementation for E and M steps are discussed in the following.

The E step Here we provide the detailed process of inferring the posterior distribution of the latent variables $\mathbf{h}_{1:S}, \mathbf{z}_{1:S,1:M}$ given the parameters of the model θ^{old} . Actually, we only infer some marginal posteriors instead of the joint posterior $p(\mathbf{z}_{s,1:M}, \mathbf{h}_s | \mathbf{X}_{s,1:M}, \mathbf{y}_s, \theta^{old})$. This is because only these marginal posteriors will be used in the following M-step. We define the following notations for these marginal posteriors γ and ξ and auxiliary variables α and β ($1 \leq s \leq S, 1 \leq m \leq M, 1 \leq n \leq N, 2 \leq n' \leq N, 1 \leq j \leq K, 1 \leq k \leq K, 1 \leq g \leq G$):

$$\alpha_{sgmnk} = p(\mathbf{x}_{sm1}, \dots, \mathbf{x}_{smn}, z_{smnk} | h_{sg}; \theta^{old}) \quad (7)$$

$$\beta_{sgmnk} = p(\mathbf{x}_{s,m,n+1}, \dots, \mathbf{x}_{smN} | z_{smnk}, h_{sg}; \theta^{old}) \quad (8)$$

Procedure forward

input : A set of the parameters θ
output: The variables α

// Initialization;
 $\alpha_{sgm1k} = \pi_{gk} \mathcal{N}(\mathbf{x}_{sm1}; \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk})$ for all s, g, m and k ;

```

1 for  $s=1$  to  $S$  do // Induction
2   for  $g=1$  to  $G$  do
3     for  $m=1$  to  $M$  do
4       for  $n=1$  to  $N-1$  do
5         for  $k=1$  to  $K$  do
6            $\alpha_{s,g,m,n+1,k} = \sum_{j=1}^K \alpha_{sgmnj} a_{gjk} \mathcal{N}(\mathbf{x}_{s,m,n+1}; \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk})$ ;
7         end
8       end
9     end
10  end
11 end

```

Procedure backward

input : A set of the parameters θ
output: The variables β

// Initialization;
 $\beta_{sgmNk} = 1$ for all s, g, m and k ;

```

1 for  $s=1$  to  $S$  do // Induction
2   for  $g=1$  to  $G$  do
3     for  $m=1$  to  $M$  do
4       for  $n=N-1$  to  $1$  do
5         for  $j=1$  to  $K$  do
6            $\beta_{sgmnk} = \sum_{j=1}^K a_{gjk} \mathcal{N}(\mathbf{x}_{s,m,n+1}; \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk}) \beta_{s,g,m,n+1,j}$ ;
7         end
8       end
9     end
10  end
11 end

```

$$\gamma_{sgmnk} = p(z_{smnk}, h_{sg} | \mathbf{X}_{sm}, \mathbf{y}_s; \theta^{old}) \quad (9)$$

$$\xi_{s,g,m,n'-1,j,n',k} = p(z_{s,m,n'-1,j}, z_{smn'k} | h_{sg}, \mathbf{X}_{sm}, \mathbf{y}_s; \theta^{old}) \quad (10)$$

The inference algorithm is presented in Algorithm 2. Specifically, line 1 calls Procedure forward to calculate the forward variables α , while line 2 calls Procedure backward to calculate the backward variables β . Then line3-15 calculate the value of each element of the posteriors γ and ξ and the h_s^* ($1 \leq s$) on the basis of the α , β and θ^{old} . These posteriors will be used in the M-step for updating the parameters.

Algorithm 2: Estep()

```

input : An initial setting for the parameters  $\theta^{old}$ 
output: Inferred posterior distributions  $\gamma$ ,  $\xi$  and  $h_s^*$  ( $1 \leq s \leq S$ )

/* Calculation of  $\alpha$ ,  $\beta$  */
1 Call Procedure forward using  $\theta^{old}$  as input;
2 Call Procedure backward using  $\theta^{old}$  as input;
/* Calculation of  $\gamma$ ,  $\xi$  and  $h_s^*$  ( $1 \leq s \leq S$ ) */
3 for  $s=1$  to  $S$  do
4   for  $g=1$  to  $G$  do
5     for  $m=1$  to  $M$  do
6        $p(\mathbf{X}_{sm}|h_{sg}; \theta^{old}) = \sum_{k=1}^K \alpha_{sgmNk}$ ;
7       for  $n=1$  to  $N$  do
8          $\gamma_{sgmnk} = \frac{\alpha_{sgmnk} \beta_{sgmnk}}{p(\mathbf{X}_{sm}|h_{sg}; \theta^{old})}$ ;
9          $\xi_{s,g,m,n-1,j,n,k} = \frac{\alpha_{s,g,m,n-1,k} \mathcal{N}(\mathbf{x}_{smn}; \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk}) a_{gjk} \beta_{sgmnk}}{p(\mathbf{X}_{sm}|h_{sg}; \theta^{old})}$  ( $n > 2$ );
10        end
11      end
12    end
13     $p(h_{sg}|\mathbf{y}_s; \theta^{old}) = \prod_{m=1}^M p(\mathbf{X}_{sm}|h_{sg}; \theta^{old})$ 
14     $p(h_{sg}|\mathbf{X}_{s,1:M}, \mathbf{y}_s; \theta^{old}) = \frac{p(h_{sg}|\mathbf{y}_s; \theta^{old}) p(h_{sg}|\mathbf{X}_{s,1:M}; \theta^{old})}{\sum_{g=1}^G p(h_{sg}|\mathbf{y}_s; \theta^{old}) p(h_{sg}|\mathbf{X}_{s,1:M}; \theta^{old})}$ ;
15     $h_s^* = \arg \max_g p(h_{sg}|\mathbf{X}_{s,1:M}, \mathbf{y}_s; \theta^{old})$ ;
15 end

```

The M step Here we provide the detailed process for M step. Basically, it updates the parameters by maximizing the $Q(\theta, \theta^{old})$ with respect to them. If substituting the distributions with inferred marginal posteriors in the Q function, we can obtain

$$\begin{aligned}
 Q(\theta, \theta^{old}) = & \sum_{s, \mathbf{h}, \mathbf{z}_{s,1:M}} p(\mathbf{z}_{s,1:M}, \mathbf{h} | \mathbf{X}_{s,1:M}, \mathbf{y}_s; \theta^{old}) \sum_{g=1}^G h_{sg} \log d_g(\mathbf{y}) \\
 & + \sum_{s,g,m,k} \gamma_{sgm1k} \log \pi_{gk} \\
 & + \sum_{s,g,m,j,k} \sum_{n=2}^N \xi_{s,g,m,n-1,j,n,k} \log a_{gjk} \\
 & + \sum_{s,g,m,n,k} \gamma_{sgmnk} \log \mathcal{N}(\mathbf{x}_{smn}; \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk}) \tag{11}
 \end{aligned}$$

Then the update formula for parameters $\mathbf{d}, \boldsymbol{\pi}, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\sigma}$ can be obtained by maximizing the Q with respect to them, respectively:

- Updating of $\mathbf{d}(\cdot)$: Maximizing Q with respect to \mathbf{d} is equivalent to maximizing the first item of Equation 11. However, \mathbf{y} is a mixture of discrete and

Procedure estimatePET

input : The data tuple (\mathbf{y}_s, h_s^*) ($1 \leq s \leq S$)
output: The learned PET $\mathbf{d}(\cdot)$

```

1 while stopping rule is not satisfactory do
2   | Examine all possible binary splits on every attribute of  $\mathbf{y}_s$  ( $1 \leq s \leq S$ );
3   | Select a split with best optimization criterion;
4   | Impose the split on the PET  $\mathbf{d}(\cdot)$ ;
5   | Repeat recursively for the two child nodes;
6 end
7 for node in the PET  $\mathbf{d}(\cdot)$  do
8   | Do Laplace correction on each node;
9 end

```

Algorithm 3: Mstep()

input : Inferred posterior distributions γ, ξ and h_s^* ($1 \leq s \leq S$)
output: The updated parameters θ^{new}

```

1 Call Procedure estimatePET to update  $\mathbf{d}(\cdot)$ ;
2 Update  $\pi, \mathbf{A}, \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk}$  according to Equation 12-15;

```

continuous values and it is impractical to find an analytic solution to \mathbf{d} . Here we consider a heuristic solution through the formation of probability estimation trees (PETs), which is a decision tree [7] with a Laplace estimation [10] of the probability on class memberships [9]. The heuristic algorithm for estimating the PET is described in Procedure estimatePET.

- Updating of $\pi, \mathbf{A}, \boldsymbol{\mu}$ and $\boldsymbol{\sigma}$: Maximization Q with respect to $\pi, \mathbf{A}, \boldsymbol{\mu}, \boldsymbol{\sigma}$ is easily achieved using appropriate Lagrange multipliers, respectively. The results are as follows:

$$\pi_{gk} = \frac{\sum_{s,m,g} \gamma_{sgm1k}}{\sum_{s,m,k,g} \gamma_{sgm1k}} \quad (12)$$

$$a_{gjk} = \frac{\sum_{s,m,n,g} \xi_{s,g,m,n-1,j,n,k}}{\sum_{s,m,l,n,g} \xi_{s,g,m,n-1,j,n,k}} \quad (13)$$

$$\boldsymbol{\mu}_{gk} = \frac{\sum_{s,m,g,n} \gamma_{sgmnk} \mathbf{x}_{smn}}{\sum_{s,m,n,g} \gamma_{sgmnk}} \quad (14)$$

$$\boldsymbol{\sigma}_{gk} = \frac{\sum_{s,m,g,n} \gamma_{sgmnk} (\mathbf{x}_{smn} - \boldsymbol{\mu}_{gk})(\mathbf{x}_{smn} - \boldsymbol{\mu}_{gk})^T}{\sum_{s,m,n,g} \gamma_{sgmnk}} \quad (15)$$

Algorithm 3 summarizes the whole process of the M step.

5 Empirical Study

The aim of the empirical study is to illustrate three issues:

- The feasibility of the learning algorithm for the CSDM. Since we have proposed an iterative (i.e., EM-based) learning method, it is pivotal to show its convergence on the gait data set.
- The predictability of the CSDM. The aim of the CSDM is to discover the correlated relationship between the static and dynamic data. Thus, it is interesting to validate its predictive power on other data falling outside the scope of the training data set.
- The usability of the CSDM. Because the CSDM is designed to be used by gait experts, we need to demonstrate intuitive knowledge extracted by the CSDM.

5.1 Experimental Settings

In this section, we provide the details of the synthetic and real-world data used in this paper. The synthetic data we sample from the true parameters listed in Table 2. We vary the s_0 for different sample sizes (e.g., $s_0 = 100, 500, 1500$) to represent relatively small, medium and large data sets. The real-world data set we use for testing is provided by the gait lab at the Royal Children’s Hospital, Melbourne⁸. We have collected the static and dynamic data for 99 patients. The static data consists of 8 attributes and is summarized in Table 3. There are at most 6 gait trials for each subject and each gait trial has 101 vector observations. In general, all the curves for both left and right sides need to be considered collectively. However, for simplicity and consistency, we only use the right side curve of the hip rotation dimension for analysis at this early pilot stage, as suggested by the gait analysis expert at the Royal Children’s Hospital, Melbourne.

5.2 Experimental Results

Convergence of the Learning Process In this section, we evaluate the convergence behavior of the learning algorithm for the CSDM. For each iteration, we calculate the averaged log-likelihood as $\frac{1}{S} \sum_{s=1}^S \sum_{m=1}^M \log p(\mathbf{X}_{sm}, \mathbf{y}_s; \boldsymbol{\theta}^{old})$, where $\boldsymbol{\theta}^{old}$ is the parameters updated from last iteration. Figure 3(a) shows the CSDM against the iteration numbers for different sample sizes of the synthetic data and Figure 3(b) shows the results of the averaged log-likelihoods for CSDMs using different numbers (represented as G) of hidden gait patterns. As expected, the averaged log-likelihood is not monotonic all the time, since part of the learning process uses a heuristic algorithm. However, the best averaged log-likelihoods are usually achieved after at most 5 iterations, which proves the convergence of the proposed learning algorithm. It can be seen from Figure 3(a),

⁸ <http://www.rch.org.au/gait/>

Table 2: The Parameters for the Synthetic Data

\mathbf{d}	if $-50 \leq y < -25$, $p(h_1 = 1 y) = 1$, if $-25 \leq y < 0$, $p(h_2 = 1 y) = 1$, if $0 \leq y < 25$, $p(h_1 = 1 y) = 1$, if $25 \leq y < 50$, $p(h_2 = 1 y) = 1$.			
$\boldsymbol{\pi}$	$\pi_{1,1:2} =$	$\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$	$\pi_{2,1:2} =$	$\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$
\mathbf{A}	$a_{1,1:2,1:2} =$	$\begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}$	$a_{2,1:2,1:2} =$	$\begin{bmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{bmatrix}$
$\boldsymbol{\mu}$	$\mu_{1,1:2,1} =$	$\begin{bmatrix} 0 \\ 3 \end{bmatrix}$	$\mu_{2,1:2,1} =$	$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$
$\boldsymbol{\sigma}$	$\sigma_{1,1:2,1} =$	$\begin{bmatrix} 1 & 1 \end{bmatrix}$	$\sigma_{2,1:2,1} =$	$\begin{bmatrix} 1 & 1 \end{bmatrix}$

Table 3: Description of the Static Data

Name of Attributes	Data Type	Value Range
internalrotation_r	continuous	23 to 90
internalrotation_l	continuous	20 to 94
externalrotation_r	continuous	-5 to 57
externalrotation_l	continuous	-26 to 51
anteversion_r	continuous	10 to 50
anteversion_l	continuous	4 to 45
hipabductors_r	categorical	$\{-1, 1, 1 + (1.5), \dots, 5\}$
hipabductors_l	categorical	$\{-1, 1, 1 + (1.5), \dots, 5\}$

a larger sample size will lead to a higher log-likelihood for the learning algorithm. For the real-world data set, $G = 4^9$ shows the fastest convergence rate of the three settings for CSDMs.

Predictive Performance To evaluate the performance of the CSDM, we measure its predictive accuracy in terms of how well the future gait profile can be predicted given the static profile and learned parameters. Since the final prediction is a set of complex variables, we measure the predictive log-likelihood $\sum_{s'=1}^{S'} \log p(\mathbf{X}_{s',1:M} | \mathbf{y}_{s'}; \boldsymbol{\theta})$ in the testing data with S' static and gait profiles, where $\boldsymbol{\theta}$ is learned from the training data. Then, the following can be obtained by using Bayes rule:

$$\log p(\mathbf{X}_{s',1:M} | \mathbf{y}_{s'}; \boldsymbol{\theta}) = \log \left(\sum_g p(h_{s'_g} | \mathbf{y}_{s'}; \boldsymbol{\theta}) p(\mathbf{X}_{s',1:M} | h_{s'_g}; \boldsymbol{\theta}) \right) \quad (16)$$

where $p(h_{s'_g} | \mathbf{y}_{s'}; \boldsymbol{\theta})$ and $p(\mathbf{X}_{s',1:M} | h_{s'_g}; \boldsymbol{\theta})$ can be calculated by using the line 13 and 14 of Algorithm 2 (i.e., E step).

⁹ The number of G is suggested by gait experts not exceeding 4.

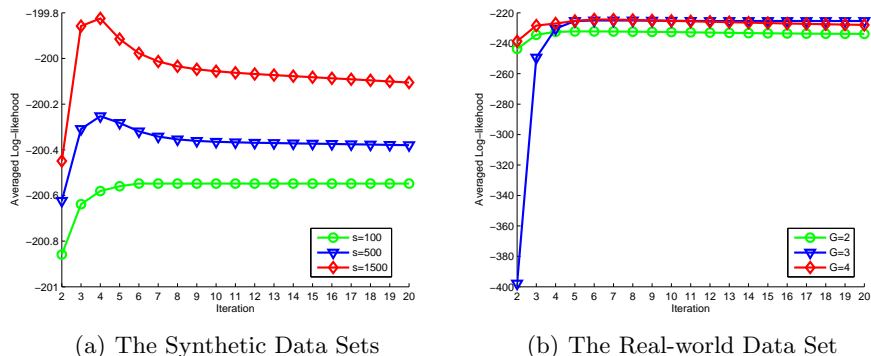


Figure 3: Log-likelihood for the CSDM against the iteration numbers for different numbers of hidden gait pattern G .

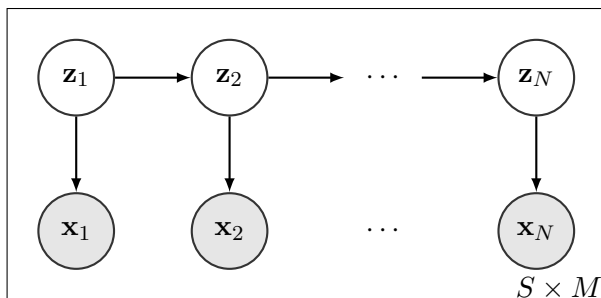


Figure 4: The Graphical Model for the Baseline Algorithm.

Without loss of generality, we propose a baseline algorithm to compare with our proposed method, which ignores the static data for modeling and prediction. The baseline model is a standard HMM with multiple observation sequences, whose graphical model is depicted in Figure 4. As indicated by the baseline model, it assumes all the gait trials are independently generated from an HMM. Using the standard algorithm provided in [1, 11], we can learn the parameters of the baseline model, denoted as θ_0 from the training data. Accordingly, the predictive averaged log-likelihood for new gait trials can be calculated as $\sum_{s'=1}^{S'} \log p(\mathbf{X}_{s',1:M}; \theta_0)$.

We compare the CSDM with the alternating baseline scheme, an HMM with multiple sequences. We report on averages over 10 times 5-fold cross validations for the synthetic and real-world data, respectively. As shown in Table 4(a), all the CSDMs with different numbers of sample sizes outperform the baseline algorithm significantly. This may be because the proposed CSDM captures the correlated relationships existed in the data rather than ignoring them. Similarly, it can be

Table 4: The Comparison of the Log-likelihoods

(a) The Synthetic Data				(b) The Real-world Data			
	$s_0 = 100$	$s_0 = 500$	$s_0 = 1500$		$G = 2$	$G = 3$	$G = 4$
CSDM	-8016	-40090	-120310	CSDM	-1310	-1388	-1299
Baseline	-8025	-40132	-120420	Baseline	-1426	-1502	-1426

observed from Table 4(b) that all the CSDMs achieve higher log-likelihoods than their counterparts of the baseline model, which proves the predictive power of our proposed CSDM on the real-world data.

Extracting Knowledge from the CSDM In this section, we provide an illustrative example of extracting intuitive knowledge from a CSDM on the gait data. Our real-world data are described in Section 5.1. We used the EM algorithm described in Section 4.4 to find the model parameters for a 4-hidden-gait-pattern CSDM as suggested by gait experts. Given the learned CSDM, we can extract the intuitive knowledge from the data set to answer the following questions:

- What kinds of static data will lead to what kinds of hidden gait patterns?
- What does the gait look like for each hidden gait pattern?

The first question is actually asking what is $p(\mathbf{h}|\mathbf{y};\boldsymbol{\theta})$ (and we omit the subscript s since all s share the same parameters). Figure 5(a) shows an answer to the first question in the form of a decision tree representation. This tree¹⁰ decides hidden gait patterns based on the 8 features of the static data (e.g., *internalrotation_r*, *externalrotation_r* and *anteversion_r*) used in the data set. To decide the hidden gait patterns based on the static data, start at the top node, represented by a triangle (\triangle). The first decision is whether *anteversion_r* is smaller than 22.5. If so, follow the left branch, and see that the tree classifies the data as gait pattern 2. If, however, anteversion exceeds 22.5, then follow the right branch to the lower-right triangle node. Here the tree asks if *hipabductors_l* is in (-1 3 3.5 4.5). If not, then follow the right branch to see that the tree classifies the data as gait pattern 4. For other nodes, the gait patterns can be decided in similar manners.

The second question is actually asking $\arg \max_g p(h_{sg}|\mathbf{X}_{s,1:M}, \mathbf{y}_s; \boldsymbol{\theta})$ ($1 \leq s \leq S$). In other words, we need to infer which gait trials belong to the corresponding hidden gait patterns in the corpus. We use line 14 described in Algorithm 2 to obtain the hidden gait pattern names of the gait trials. We can then plot representative gaits for each hidden gait pattern to answer the second question above, as shown in Figures 5(b)-5(e). Figure 5(b) shows a collection of gaits for

¹⁰ For simplicity, we do not display the fully tree and only display the gait pattern with the highest probability.

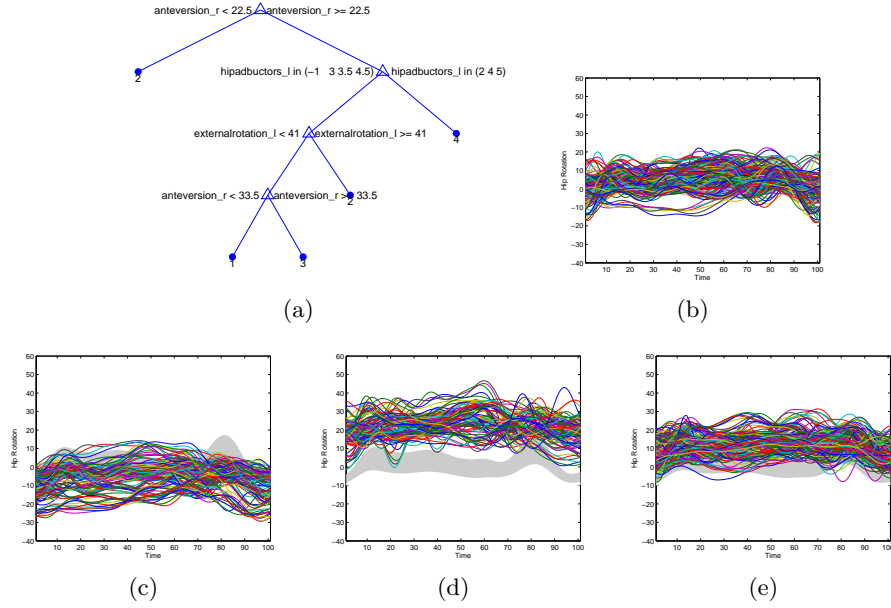


Figure 5: Extracted Knowledge from the CSDM: (a) The Decision Tree to Predict Gait Patterns Given the Static Data, (b)-(e) Representative Gaits for Gait Pattern 1-4.

the hidden gait pattern 1. We can see that most of them fall into the normal area, which may indicate that these gaits are good. Figure 5(e) shows a collection of gaits for the hidden gait pattern 4 and most of them are a little below the normal area, indicating that these gaits are not as good. By contrast, most of the gaits in Figure 5(d) representing hidden gait pattern 3 fall outside the normal area and are abnormal gaits. Figure 5(c) shows that the representative gaits for hidden gait pattern 2 are below the normal area but that there are only a few. This indicates these gaits are only slightly abnormal.

6 Conclusions and Future Work

This paper presents a new probabilistic graphical model (i.e., CSDM) for quantitatively discovering the correlated relationship between static physical examination data and dynamic gait data in clinical gait analysis. To learn the parameters of the CSDM given the training data, we proposed an EM-based algorithm. One of the main advantages of the CSDM is its ability to provide intuitive knowledge. For example, the CSDM informs us what kinds of static data will lead to what kinds of hidden gait patterns and what the gaits look like for each hid-

den gait pattern. The experiments on both synthetic and real-world data (excerpted from patient records at the Royal Children’s Hospital, Melbourne) show promising results in terms of learning convergence, predictive performance and knowledge discovery. One direction for future work is to improve the CSDM with semi-supervised learning. Currently the CSDM is learned totally unsupervised, which may generate unexpected results due to its highly stochastic nature. Further collaboration with gait analysis experts may alleviate this problem through the manual labeling some examples. In addition, to enhance the learning of the model, we plan to collect more real-world data and include more dimensions of the dynamic data for analysis.

Acknowledgments.

To be added for the camera-ready version if accepted.

References

1. Baum, L., Petrie, T., Soules, G., Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics* 41(1), 164–171 (1970)
2. Bishop, C.: *Pattern recognition and machine learning*. Information Science and Statistics, Springer, New York (2006)
3. Chau, T.: A review of analytical techniques for gait data. part 1: fuzzy, statistical and fractal methods. *Gait & posture* 13(1), 49–66 (2001)
4. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1), 1–38 (1977)
5. Desloovere, K., Molenaers, G., Feys, H., Huenaeerts, C., Callewaert, B., Walle, P.: Do dynamic and static clinical measurements correlate with gait analysis parameters in children with cerebral palsy? *Gait & Posture* 24(3), 302–313 (2006)
6. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the Eighteenth International Conference on Machine Learning*. pp. 282–289. Morgan Kaufmann Publishers Inc. (2001)
7. Olshen, L., Stone, C.: *Classification and regression trees*. Wadsworth International Group (1984)
8. Perry, J., Davids, J.: Gait analysis: normal and pathological function. *Journal of Pediatric Orthopaedics* 12(6), 815 (1992)
9. Provost, F., Domingos, P.: Tree induction for probability-based ranking. *Machine learning* 52(3), 199–215 (2003)
10. Provost, F., Fawcett, T.: Robust classification for imprecise environments. *Machine learning* 42(3), 203–231 (2001)
11. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. *Readings in speech recognition* 53(3), 267–296 (1990)
12. Sagawa, Y., Watelain, E., De Coulon, G., Kaelin, A., Armand, S.: What are the most important clinical measurements affecting gait in patients with cerebral palsy? *Gait & Posture* 36, S11–S12 (2012)
13. Zhang, B.L., Zhang, Y., Begg, R.K.: Gait classification in children with cerebral palsy by bayesian approach. *Pattern Recognition* 42(4), 581–586 (2009)