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

Yin Song[1], Jian Zhang[1], Longbing Cao[1], and Morgan Sangeux[234]

[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. Patients may have an abnormal gait due to a range of physical impairment or brain damage. Clinical gait analysis (CGA) is a technique for identifying the underlying impairments that affect a patient's gait pattern. The CGA is critical for 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. This process is carried out by gait analysis experts, mainly based on their experience which may lead to subjective diagnoses. To facilitate the automation of this process and form a relatively objective diagnosis, this paper proposes a new probabilistic correlated static-dynamic model (CSDM) to discover correlated relationships between the dynamic characteristics of 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 dynamic characteristics based on static data. Our initial experiments indicate that the CSDM is promising for discovering the correlated relationship between 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

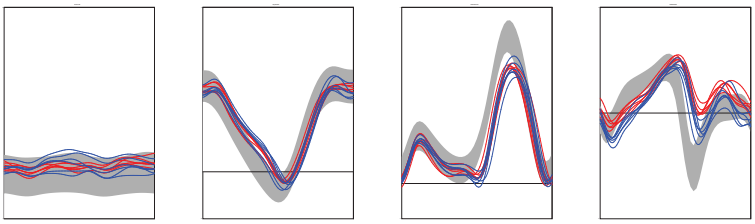
Table 1: An Excerpt Data Set from the Static Data

| 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 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

determine a patient's impairments to plan manageable treatment. Usually, two types of data are used in clinical gait analysis: *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 abductor muscles. 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 gait curve examples for one subject. Gait curves 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, there are multiple curves for each dimension. In addition, each dimension has both the left and right side of the body. 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* 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.

In practice, static data is used to explain abnormal features in dynamic data. In other words, gait analysis experts try to discover the correlated relationships between *static* and *dynamic* data for further clinical diagnosis. This process has been conducted empirically by clinical experts and thus is *qualitative*. In this paper, we make an initial exploration to discover the *quantitative* correlated relationships between the *static* data and *dynamic* curves.

The rest of the paper is organize as following: The next section reviews the work related to this paper and Section 3 follows by the problem formalization. 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. Experimental results on both synthetic and real-world data sets are reported in Section 5 and Section 6 concludes this paper.



(a) (b) (c) (d)

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.

## 2 Related Work

Recent research in CGA [3, 5, 13, 12] have made initial attempts at the automatic discovery of correlated relationships in clinical gait data by machine learning methods such as multiple linear regression [5] and fuzzy decision trees [12]. However, previous researchers usually preprocessed the gait data and discarded the dynamic characteristics of that data, which fails to explore the correlated relationship between static data and dynamic curves. To the best of our knowledge, our work is a first attempt to explore this correlated relationship comprehensively.

Probabilistic models related to this paper exists, for example, hidden Markov models (HMMs) [11] and conditional random fields (CRFs) [6]. Since these models focus on modeling dynamic curves, 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

The following terms are defined:

- A *static profile* is a collection of static physical examination features of one subject denoted by $\mathbf{y} = (y_1, y_2, \cdots, y_L)$, 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 $\mathbf{X}_{1:M} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_M\}$.
- A *gait trial* (cycle) is multivariate time series denoted by $\mathbf{X}_m = (\mathbf{x}_{m1}, \mathbf{x}_{m2}, \cdots, \mathbf{x}_{mN})$, 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 $\mathbf{x}_{mj} = \left[x_{m1j}\ x_{m2j} \cdots x_{mDj}\right]^T$

($D$ 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 $(x_{mj1}, x_{mj2}, \cdots, x_{mjN})$ $(1 \leq j \leq D)$ 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 was to develop a probabilistic model $p(\mathbf{X}_{1:M}, \mathbf{y})$ that considers the correlated relationships between the *static profile* (i.e., static data) and the corresponding *gait profile* (i.e., 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, *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 $p(\mathbf{X}_{1:M}|\mathbf{h})$, the generating process of it is not intuitive. Thus, 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

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 *gait* profiles $p(\mathbf{X}_{1:M}, \mathbf{y})$. The graphical model for the CSDM is shown in Figure 2 (subscript $m$ is omitted for convenience). We use conventional notation to represent the graphical model [2]. In Figure 2,

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

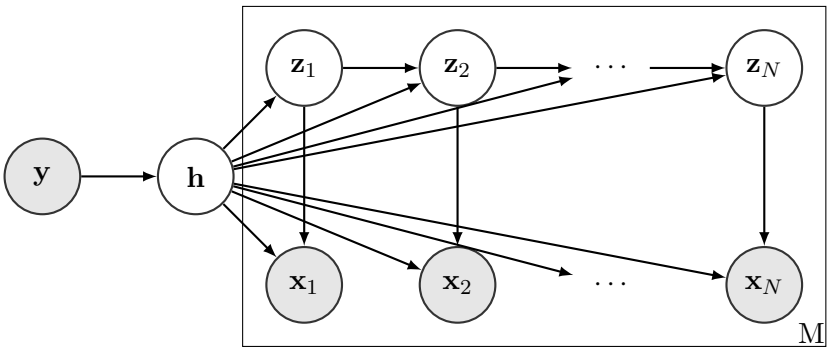



Figure 2: The Graphical Model of the CSDM

each node represents a random variable (or group of random variables). For instance, a *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 profile* $\mathbf{y}$ 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 profile* $\mathbf{y}$ 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 following[6]:

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

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

where $d_g$ $(1 \leq g \leq G)$ is a set of mapping functions $(\mathbf{y} \rightarrow d_g(\mathbf{y}) \equiv p(h_g = 1|\mathbf{y}))$ 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 $p(\mathbf{h}|\mathbf{y};\mathbf{d})$. 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 $d_g(\mathbf{y})$".

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

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

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

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

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

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

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 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 is 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].

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

[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 $\boldsymbol{\theta}^{old}$
**Output**: Learned parameters $\boldsymbol{\theta}^{new}$

```
while the convergence criterion is not satisfied do
    Estep();
    θ^new = Mstep();
end
```

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

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

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

**Challenges of the Learning Algorithms** The challenges of the above algorithm is in 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 $\mathbf{h}_s$ $(1 \leq s \leq S)$. We provided 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 was difficult to find an analytic solution for $\mathbf{d}(\cdot)$. We utilized 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 $\boldsymbol{\theta}^{old}$. Actually, we only infer some marginal posteriors instead of the joint posterior $p(\mathbf{z}_{s,1:M}, \mathbf{h}_s|\mathbf{X}_{s,1:M}, \mathbf{y}_s, \boldsymbol{\theta}^{old})$. This is because only these marginal posteriors will be used in the following M-step. We define the following notations for these marginal posteriors $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ and auxiliary variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ $(1 \leq s \leq S, 1 \leq m \leq M, 1 \leq n \leq N, 2 \leq n^{'} \leq N, 1 \leq j \leq K, 1 \leq k \leq K, 1 \leq g \leq G)$:

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

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

**Procedure** forward

**input** : A set of the parameters $\boldsymbol{\theta}$
**output**: The variables $\boldsymbol{\alpha}$

// Initialization;
$\alpha_{sgm1k} = \pi_{gk}\mathcal{N}(\mathbf{x}_{sm1}; \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk})$ for all $s$, $g$, $m$ and $k$;
**for** *s=1* **to** $S$ **do** // Induction
  **for** *g=1* **to** $G$ **do**
    **for** *m=1* **to** $M$ **do**
      **for** *n=1* **to** *N-1* **do**
        **for** *k=1* **to** $K$ **do**
          $\alpha_{s,g,m,n+1,k} = \sum_{j=1}^{K} \alpha_{sgmnj} a_{gjk} \mathcal{N}(\mathbf{x}_{s,m,n+1}; \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk})$;
        **end**
      **end**
    **end**
  **end**
**end**

**Procedure** backward

**input** : A set of the parameters $\boldsymbol{\theta}$
**output**: The variables $\boldsymbol{\beta}$

// Initialization;
$\beta_{sgmNk} = 1$ for all $s$, $g$, $m$ and $k$;
**for** *s=1* **to** $S$ **do** // Induction
  **for** *g=1* **to** $G$ **do**
    **for** *m=1* **to** $M$ **do**
      **for** *n=N-1* **to** *1* **do**
        **for** *j=1* **to** $K$ **do**
          $\beta_{sgmnk} = \sum_{j=1}^{K} a_{gjk} \mathcal{N}(\mathbf{x}_{s,m,n+1}; \boldsymbol{\mu}_{gk}, \boldsymbol{\sigma}_{gk}) \beta_{s,g,m,n+1,j}$;
        **end**
      **end**
    **end**
  **end**
**end**

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

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

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

**Algorithm 2:** Estep()

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

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

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

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 $\mathbf{d}$: 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 con-

**Procedure** estimatePET

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

```
while stopping rule is not satisfactory do
    Examine all possible binary splits on every attribute of y_s (1 ≤ s ≤ S);
    Select a split with best optimization criterion;
    Impose the split on the PET d;
    Repeat recursively for the two child nodes;
end
for node in the PET d(·) do
    Do Laplace correction on each node;
end
```

**Algorithm 3:** Mstep()

**input** : Inferred posterior distributions $\boldsymbol{\gamma}$, $\boldsymbol{\xi}$ and $h_s^*$ $(1 \leq s \leq S)$
**output**: The updated parameters $\theta^{new}$

```
Call Procedure estimatePET to update d(·);
Update π, A, μ_gk, σ_gk according to Equation 12-15;
```

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

$$\pi_{gk} = \frac{\sum\limits_{s,m,g} \gamma_{sgm1k}}{\sum\limits_{s,m,k,g} \gamma_{sgm1k}} \tag{12}$$

$$a_{gjk} = \frac{\sum\limits_{s,m,n,g} \xi_{s,g,m,n-1,j,n,k}}{\sum_{s,m,l,n,g} \xi_{s,g,m,n-1,j,n,k}} \tag{13}$$

$$\boldsymbol{\mu}_{gk} = \frac{\sum\limits_{s,m,g,n} \gamma_{sgmnk} \mathbf{x}_{smn}}{\sum\limits_{s,m,n,g} \gamma_{sgmnk}} \tag{14}$$

$$\boldsymbol{\sigma}_{gk} = \frac{\sum\limits_{s,m,g,n} \gamma_{sgmnk} (\mathbf{x}_{smn} - \mu_{gk})(\mathbf{x}_{smn} - \mu_{gk})^T}{\sum\limits_{s,m,n,g} \gamma_{sgmnk}} \tag{15}$$

Algorithm 3 summarizes the whole process of the M step.

Table 2: The Parameters for the Synthetic Data

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

## 5 Empirical Study

The aim of this study is to test:

- 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

we sampled the synthetic data from the true parameters listed in Table 2. We varied 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 used was provided by the gait lab at the Royal Children's Hospital, Melbourne[8]. We have collected a subset of static and dynamic data for 99 patients. The static data subset consisted of 8 attributes summarized in Table 3. There were at most 6 gait trials for each subject and each gait trial had 101 vector observations. In principle, curves for both left and right sides may be included. However, for simplicity and consistency, we only used the right side curves of the hip rotation dimension for analysis in this pilot study.

[8] http://www.rch.org.au/gait/

Table 3: Description of the Static Data

| Name of Attributes | Data Type | Value Range |
|---|---|---|
| internalrotation_r (ir_r) | continuous | 23 to 90 |
| internalrotation_l (ir_l) | continuous | 20 to 94 |
| externalrotation_r (er_r) | continuous | -5 to 57 |
| externalrotation_l (er_l) | continuous | -26 to 51 |
| anteversion_r (a_r) | continuous | 10 to 50 |
| anteversion_l (a_l) | continuous | 4 to 45 |
| hipabductors_r (h_r) | discrete | -1 to 5 |
| hipabductors_l (h_l) | discrete | -1 to 5 |

### 5.2 Experimental Results

**Convergence of the Learning Process** 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), 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.

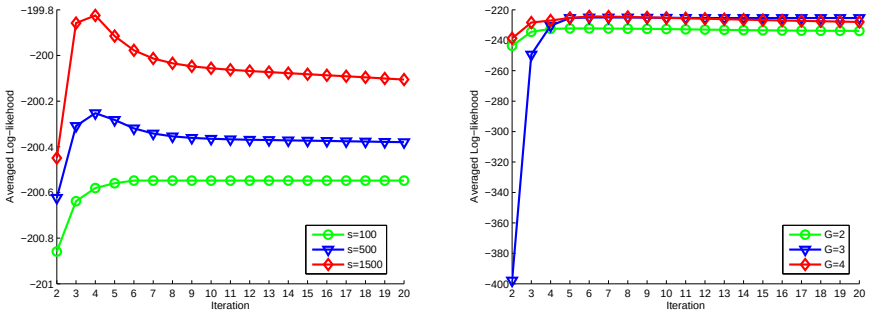
**Predictive Performance** We measured the CSDM predictive accuracy in terms of how well the future gait profile can be predicted given the static profile and learned parameters. Since the final prediction is a set of complex variables, we measure the predictive log-likelihood $\sum_{s'=1}^{S'}\log p(\mathbf{X}_{s',1:M}|\mathbf{y}_{s'};\boldsymbol{\theta})$ 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:

$$\log p(\mathbf{X}_{s',1:M}|\mathbf{y}_{s'};\boldsymbol{\theta}) = \log(\sum_{g} p(h_{s'g}|\mathbf{y}_{s'};\boldsymbol{\theta})p(\mathbf{X}_{s',1:M}|h_{s'g};\boldsymbol{\theta})) \tag{16}$$

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).

Without loss of generality, we proposed a baseline algorithm which ignored the static data for modeling and prediction to compare with our proposed

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




(a) The Synthetic Data Sets (b) The Real-world Data Set

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

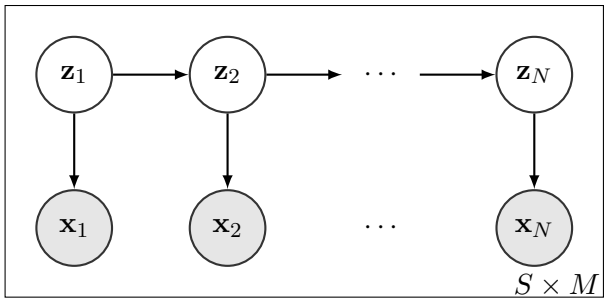



Figure 4: The Graphical Model for the Baseline Algorithm.

method. The baseline model is a standard HMM with multiple observation sequences, whose graphical model is depicted in Figure 4. 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 $\boldsymbol{\theta}_0$ from the training data. Accordingly, the predictive averaged log-likelihood for new gait trials can be calculated as $\sum_{s'=1}^{S'} \log p(\mathbf{X}_{s',1:M}; \boldsymbol{\theta}_0)$.

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 outperformed the baseline algorithm significantly. This may be because the proposed CSDM captures the correlated relationships existing in the data rather than ignoring them. Similarly, it can be observed from Table 4(b) that all the CSDMs achieved higher log-likelihoods than their counterparts of the baseline model. This proves the predictive power of our proposed CSDM on real-world data.

Table 4: The Comparison of the Log-likelihoods

(a) The Synthetic Data

| | $s_0 = 100$ | $s_0 = 500$ | $s_0 = 1500$ |
|---|---|---|---|
| CSDM | **-8016** | **-40090** | **-120310** |
| Baseline | -8025 | -40132 | -120420 |

(b) The Real-world Data

| | $G = 2$ | $G = 3$ | $G = 4$ |
|---|---|---|---|
| CSDM | **-1310** | **-1388** | **-1299** |
| Baseline | -1426 | -1502 | -1426 |

**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 subscript $s$ is omitted 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 tree[10] decides hidden gait patterns based on the 8 features of the static data (e.g., $ir\_r$, $er\_r$ and $a\_r$) 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 $ir\_r$ is smaller than 57. If so, follow the left branch, and see that the tree classifies the data as gait pattern 2. If, however, anteversion exceeds 57, then follow the right branch to the lower-right triangle node. Here the tree asks whether $er\_r$ is is smaller than 21.5. If so, then follow the right branch to see the question of next node until the tree classifies the data as ones of the gait patterns. 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(e) shows a collection of gaits for the hidden gait pattern 4. We can see that most of them fall into the normal area, which may indicate that these gaits are good. Figure 5(c) shows a collection of gaits for the hidden gait pattern 2 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(b) representing hidden gait pattern 1 fall outside the normal area and are abnormal gaits. Figure 5(d) shows that the representative gaits

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

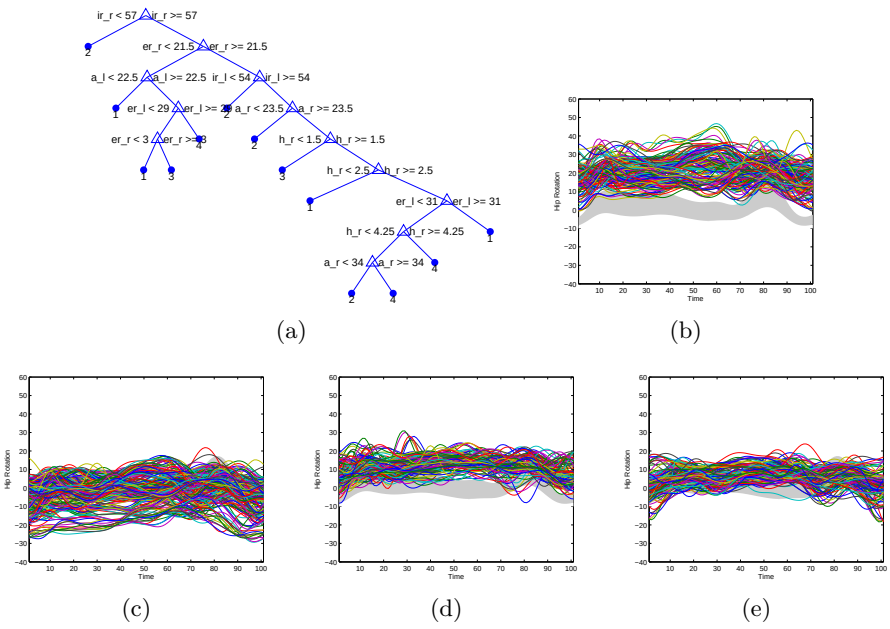



Figure 5: Extracted Knowledge from the CSDM: (a) The Decision Tree to Predict Gait Patterns Given the Static Data, (b)-(e) Representative Gaits for Gait Pattern 1-4.

for hidden gait pattern 3 are slightly above the normal area, which indicates these gaits are only slightly abnormal. Most subjects displaying pattern 1 and some subjects displaying pattern 3 would be susceptible to have surgery. By extracting the different paths that lead to those two patterns from the decision tree in Figure 5(a), we can infer what combinations of static data may have clinical implications.

## 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 on a training data set, 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 hidden gait pattern. The experiments on both synthetic and real-world data (excerpted from patient records at the Royal Children's Hospital, Melbourne) showed

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 manual labeling of some examples. We also plan to collect more real-world data and include all static and dynamic outputs from clinical gait analysis.

## Acknowledgments.

We thank the four anonymous reviewers for their valuable comments on our manuscript!

## 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., Huenaerts, 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)