Methodology Report

A Bayesian Approach to Multistage Fitting of the Variation of the Skeletal Age Features

Dong Hua,[1] Dechang Chen,[2] Fang Liu,[3] and Abdou Youssef[1]

[1] *Department of Computer Science, The George Washington University, 801 22nd Street NW, Washington, DC 20052, USA*

[2] *Division of Epidemiology and Biostatistics, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD 20814, USA*

[3] *Department of Computer Science, University of Texas-Pan American, 1201 W. University Drive, Edinburg, TX 78539, USA*

Correspondence should be addressed to Dechang Chen, dchen@usuhs.mil

Accurate assessment of skeletal maturity is important clinically. Skeletal age assessment is usually based on features encoded in ossification centers. Therefore, it is critical to design a mechanism to capture as much as possible characteristics of features. We have observed that given a feature, there exist stages of the skeletal age such that the variation pattern of the feature differs in these stages. Based on this observation, we propose a Bayesian cut fitting to describe features in response to the skeletal age. With our approach, appropriate positions for stage separation are determined automatically by a Bayesian approach, and a model is used to fit the variation of a feature within each stage. Our experimental results show that the proposed method surpasses the traditional fitting using only one line or one curve not only in the efficiency and accuracy of fitting but also in global and local feature characterization.

1. Introduction

Hand X-ray shown in Figure 1 is commonly used for skeletal age assessment in pediatric radiology. A discrepancy between skeletal maturity and the chronical age may indicate the presence of some abnormality in skeletal growth. This abnormality has been found to be related to various diseases such as endocrine disorders [1], metabolic/growth abnormalities [2], malformations and bone dysplasias [3], and gonadal dysgenesis [4]. Therefore, the assessment of skeletal maturity has become more and more important clinically. Clearly the accuracy in assessment is of the first concern.

Features encoded in ossification centers form the basis for assessment. If we know the exact characteristics of the features with regard to different stages of ages, we can do the best job on assessment. In reality, one needs a mechanism to capture such characteristics of features. Given data of a feature with respect to skeletal ages, a simple and common approach is to fit a line or a curve, which in turn is used for future prediction of new patients or assisting radiologists to understand the variation rules of the feature.

For instance, Figure 2(a) shows the variation of a ratio feature [5, 6] in vertical axis with regard to the increasing skeletal age along the horizontal axis from newborn to 19 year old boys. (More details on this ratio are provided in Section 3.2.) Here in the figure, a single line is used for fitting the values of the feature. Obviously, a line is not enough to capture the characteristic of the values of the feature. A quadratic curve, shown in Figure 2(c), does not do a good job either. Fitting a more complex curve does not seem to be a feasible approach. This is because sometimes there are available only a small amount of data which could restrict the learning of complex curves, and local properties (with respect to the time) of the feature are often lost when fitting a global complex curve, and thus leading to inaccurate future prediction.

In this paper, we propose to fit the variation of features of the skeleton age via a multistage fitting approach. With our approach, we divide the skeletal age axis into several stages or phases, and within each stage, a relative simple model (line or curve) is employed for the purpose of fitting. Usually, the variation of a feature does not follow a simple rule

Figure 1: Hand X-ray used in skeletal age assessment.

when skeletal age increases. Instead, it often shows different variation patterns among different stages of age. As shown in Figures 2(b) and 2(d), multistage fitting not only can capture the entire pattern of feature variation but also carry the local properties regarding the skeletal age. A critical question is then, how does one determine the appropriate positions to separate the stages? The proposed *Bayesian cut* in this paper provides an answer via a Bayesian approach.

The rest of the paper is organized as follows. In Section 2, we describe our models for fitting, where the Bayesian cut is introduced. In Section 3, we present our experimental results on multi-stage fitting for artificial and real data. We conclude our paper in Section 4.

2. The Proposed Method

In this section, we first describe our proposed method for a simple case and then extend it to a general scenario.

Given a sequence of values $f_1, f_2, \ldots, f_n$, which denotes the skeletal age f in an ascending order, consider the linear relationship between f and one feature y found in the hand X-ray (e.g., length of digit). Usually, such a linear relationship varies as the skeletal age increases. That is, one linear form established for one interval of the skeletal age may not hold for the next interval, where a different linear form should be used. The time where two linear forms differ is called a *change point*. Our model that takes into account linear relationships and change points is stated as follows:

$$
\begin{aligned}
y_i &= \beta_{11} + \beta_{12} f_i + \epsilon_{1i}, \quad i = 1, \ldots, t_1 \, (t_0 = 0), \\
y_i &= \beta_{21} + \beta_{22} f_i + \epsilon_{2i}, \quad i = t_1 + 1, \ldots, t_2, \\
&\vdots \\
y_i &= \beta_{k1} + \beta_{k2} f_i + \epsilon_{ki}, \quad i = t_{k-1} + 1, \ldots, t_k \, (t_k = n),
\end{aligned}
\tag{1}
$$

where $t_1, \ldots, t_{k-1}$ (correspondingly $f_1, \ldots, f_{k-1}$) indicate the sequential change points, $t_j - t_{j-1} \geq 3 \; (j = 1, \ldots, k)$, and ϵ_{ji} (for all i) are independent $N(0, \sigma_j^2)$ and ϵ_{ji} (for all i, j)

are independent of each other. In the model, the parameters β_{j1}, β_{j2}, σ_j^2, t_j are all unknown, which will be estimated in light of the given data. The interval $[t_j - t_{j-1}]$ represents the jth stage or phase, denoted by ph_j. The main task here is to estimate the times t_j. Given the estimates of t_j, the linear forms and the associated parameters can be obtained through the traditional regression technique. We note that the requirement $t_j - t_{j-1} \geq 3 \, (j = 1, \ldots, k)$ is needed for estimation of the regression lines. When $k = 2$, the model will be reduced to the two-phase regression with a single change point in [7].

The above model that uses only one dependent variable f can be generalized to include multiple independent variables. This generalization leads to the following model:

$$
\begin{aligned}
y_i &= \vec{\beta}_1^T \mathbf{f_i} + \epsilon_{1i}, \quad i = 1, \ldots, t_1 \, (t_0 = 0), \\
y_i &= \vec{\beta}_2^T \mathbf{f_i} + \epsilon_{2i}, \quad i = t_1 + 1, \ldots, t_2, \\
&\vdots \\
y_i &= \vec{\beta}_k^T \mathbf{f_i} + \epsilon_{ki}, \quad i = t_{k-1} + 1, \ldots, t_k \, (t_k = n),
\end{aligned}
\tag{2}
$$

where $\mathbf{f_i}$ is a p-dimensional vector of variables, $\vec{\beta}_j \, (j = 1, \ldots, k)$ is a p-dimensional vector of parameters, $t_j - t_{j-1} \geq p + 1$, and ϵ_{ji} are as the same as before. We refer p as the cardinality of the input vector $\mathbf{f_i}$, denoted by $C(\mathbf{f_i})$, and the number of sample points in ph_j as the cardinality of $[t_j - t_{j-1}]$, denoted by $C(ph_j)$. We note that though linear regression is used for each phase in model (2), this model certainly encompasses other nonlinear cases such as polynomial forms.

We now describe a Bayesian approach to estimate the change points. Denote $(\mathbf{f_{t_{j-1}+1}}, \ldots, \mathbf{f_{t_j}})^T$ by F_j, $(F_1^T, \ldots, F_k^T)^T$ by F, $(y_{t_{j-1}+1}, \ldots, y_{t_j})^T$ by $\mathbf{y_j}$, $(\mathbf{y_1}^T, \ldots, \mathbf{y_k}^T)^T$ by $\mathbf{y}$, and $(t_1, \ldots, t_{k-1})$ by $\mathbf{t}$. For simplicity, we assume the noninformative or uniform prior for $\vec{\beta}_j \, (j = 1, \ldots, k)$, $\ln(\sigma_j^2)$ and $\mathbf{t}$. Noninformative priors are used when information about parameters is completely unknown or when proper priors such as conjugate priors do not apply. (For a vigorous discussion on the choice of priors, see [8].) We can show the following main result (see the Appendix). Given the data $\mathbf{y}$ and the uniform prior for $\vec{\beta}_j \, (j = 1, \ldots, k)$, $\ln(\sigma_j^2)$ and $\mathbf{t}$, where the number k is predetermined, the posterior probability that change points occur at $\mathbf{t}$ is

$$
\begin{aligned}
p(\mathbf{t} \mid \mathbf{y}) = J 2^{(n-kp)/2} \prod_j \left| F_j^T F_j \right|^{-1/2} \\
\times \Gamma\left(\frac{t_j - t_{j-1} - p}{2} \right) S_j^{-(t_j - t_{j-1} - p)/2},
\end{aligned}
\tag{3}
$$

where $J = \left(\sum_{\mathbf{t}} 2^{(n-kp)/2} \prod_j |F_j^T F_j|^{-1/2} \Gamma((t_j - t_{j-1} - p)/2) \times S_j^{-(t_j - t_{j-1} - p)/2} \right)^{-1}$, and $S_j = (\mathbf{y_j} - F_j \hat{\vec{\beta}}_j)^T (\mathbf{y_j} - F_j \hat{\vec{\beta}}_j)$ with $\hat{\vec{\beta}}_j = (F_j^T F_j)^{-1} F_j^T \mathbf{y_j}$ denoting the least-squares estimator of $\vec{\beta}_j$. Using this result, we estimate $\mathbf{t}$ by t^* at which $p(\mathbf{t} \mid \mathbf{y})$
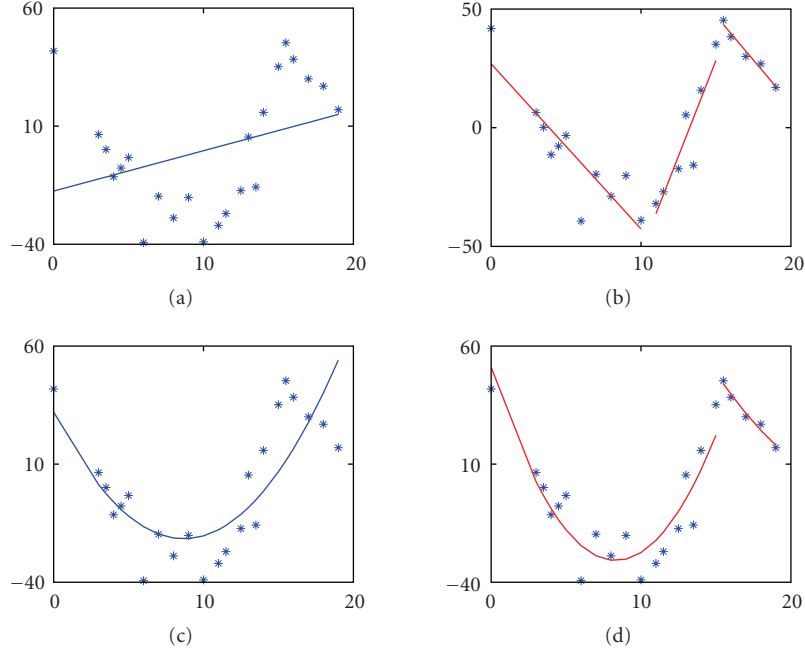
(a)

(b)

(c)

(d)

FIGURE 2: Examples of fitting the variation of the ratio feature. The horizontal axis represents the skeletal age and the vertical axis corresponds to the values of the feature.

TABLE 1: Models for testing the performance of the Bayesian cut.

m_1	$y_i = \beta_{j1} + \beta_{j2} f_i + \epsilon_{ji},$
	$\mathbf{t} = (t_1, \ldots, t_{k-1})$
m_2	$y_i = \beta_{j1} + \beta_{j2} f_i + \beta_{j3} f_i^2 + \epsilon_{ji},$
	$\mathbf{t} = (t_1, \ldots, t_{k-1})$
m_3	$y_i = \beta_{j1} + \beta_{j2} f_i + \beta_{j3} f_i^2 + \beta_{j4} f_i^3 + \epsilon_{ji},$
	$\mathbf{t} = (t_1, \ldots, t_{k-1})$
m_4	$y_i = \beta_{j1} + \beta_{j2} f_i + \beta_{j3} f_i^2 + \beta_{j4} f_i^3 + \beta_{j5} f_i^4 + \epsilon_{ji},$
	$\mathbf{t} = (t_1, \ldots, t_{k-1})$
m_5	$y_i = \beta_{j1} + \beta_{j2} f_i + \beta_{j3} f_i^2 + \beta_{j4} f_i^3 + \beta_{j5} f_i^4 + \beta_{j6} f_i^5 + \epsilon_{ji},$
	$\mathbf{t} = (t_1, \ldots, t_{k-1})$

TABLE 2: Experimental setting.

β_{ji}	$(-5.0, 5.0)$
ϵ_{ji}	$\sim N(0, \sigma_j^2),\ \sigma_j^2 \in (0, 5^{C(\mathbf{f_i})-1})$
k	$2,\ 3,\ 4$
$C(ph_j)$	$(C(\mathbf{f_i})+1), \ldots, (C(\mathbf{f_i})+1)+s$
$scale$	$1, \ldots, 10$
t_0	0
t_j	$t_{j-1} + C(ph_{j-1})$
f_i	$1, \ldots, t_k$

has its maximum, that is, $t^* = \arg\max_t p(\mathbf{t} \mid \mathbf{y})$. We call t^* the *Bayesian cut*, and the value $2^{(n-kp)/2}\prod_j |F_j^T F_j|^{-1/2}\Gamma((t_j - t_{j-1} - p)/2)S_j^{-(t_j - t_{j-1} - p)/2}$ the *proportional posterior* (pp).
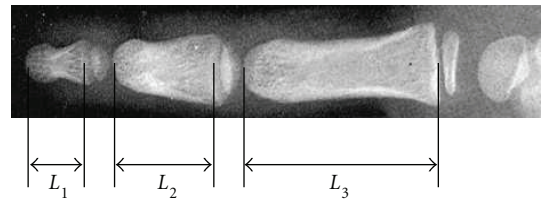
FIGURE 3: Illustration Of L_1, L_2 and L_3.

3. Experiments

In this section, we perform the Bayesian cut on two data sets: one is synthesized and the other is real. We use the synthesized data for performance evaluation in terms of recovery of changing points. The real data are used to discover the Bayesian cut and describe the feature in a multistage way which has more accurate prediction of the skeletal age compared with fitting by a single line or curve. Both linear and nonlinear regression are used for comparison. For convenience, we call the fitting with a single line or curve the *single fitting* and the fitting with the Bayesian cut the *Bayesian cut fitting*.

3.1. Synthesized Data. We consider five cases or models describing the relationship between the dependent and independent variables. These are shown in Table 1 where the input vector $\mathbf{f_i}$ for models m_1, m_2, m_3, m_4, and m_5 is $(1, f_i)^T$, $(1, f_i, f_i^2)^T$, $(1, f_i, f_i^2, f_i^3)^T$, $(1, f_i, f_i^2, f_i^3, f_i^4)^T$, and $(1, f_i, f_i^2, f_i^3, f_i^4, f_i^5)^T$, respectively. The data are generated according to the setting given in Table 2. Specifically, β_{ji} is randomly chosen from $(-5.0, 5.0)$. ϵ_{ji} is generated from a normal distribution with mean 0 and variance σ_j^2 randomly selected from $(0, 5^{C(\mathbf{f_i})-1})$. The number of sample points of

TABLE 3: AD scores for models in Table 1.

k	s	m_1	m_2	m_3	m_4	m_5
2	1	0.280	0.340	0.320	0.080	0.180
	2	0.300	0.460	0.360	0.200	0.100
	3	0.260	0.400	0.320	0.100	0.100
	4	0.640	0.380	0.260	0.180	0.180
	5	0.480	0.680	0.480	0.100	0.060
	6	0.380	0.300	0.560	0.220	0.100
	7	0.540	0.520	0.340	0.280	0.100
	8	0.900	0.520	0.440	0.120	0.020
	9	0.740	0.340	0.080	0.040	0.020
	10	0.740	0.720	0.160	0.200	0.020
3	1	0.230	0.360	0.210	0.240	0.090
	2	0.440	0.390	0.190	0.080	0.060
	3	0.590	0.340	0.210	0.220	0.060
	4	0.820	0.590	0.260	0.060	0.010
	5	0.970	0.690	0.530	0.020	0.090
	6	0.670	0.580	0.120	0.060	0.070
	7	1.220	0.750	0.160	0.080	0.190
	8	1.260	0.680	0.650	0.040	0.030
	9	1.210	0.860	0.370	0.380	0.010
	10	1.340	0.360	0.680	0.020	0.020
4	1	0.333	0.300	0.133	0.040	0.053
	2	0.440	0.433	0.227	0.060	0.033
	3	0.867	0.480	0.113	0.080	0.033
	4	0.780	0.513	0.093	0.080	0.133
	5	1.020	0.887	0.453	0.133	0.173
	6	1.360	0.760	0.193	0.093	0.180
	7	1.007	0.593	0.353	0.047	0.040
	8	0.727	0.587	0.453	0.093	0.113
	9	1.080	1.240	0.867	0.360	0.087
	10	1.213	0.873	0.333	0.120	0.140

TABLE 4: Some features of the skeletal age.

Age (yr)	L_1/L_2	L_2/L_3	$n(L_1/L_2)$	$n(L_2/L_3)$
0	0.6795	0.7016	41.8212	51.1987
3	0.6307	0.5853	6.4071	−17.6281
3.5	0.6220	0.6298	0.1020	8.6933
4.0	0.6060	0.5993	−11.4491	−9.3140
4.5	0.6111	0.5708	−7.7721	−26.1616
5.0	0.6172	0.5070	−3.3303	−63.8970
6.0	0.5675	0.5924	−39.3612	−13.4245
7.0	0.5947	0.6626	−19.6939	28.0937
8.0	0.5820	0.6097	−28.9032	−3.1878
9.0	0.5939	0.5968	−20.2149	−10.7828
10.0	0.5680	0.6643	−39.0383	29.1323
11.0	0.5776	0.6696	−32.0541	32.2560
11.5	0.5845	0.6550	−27.0602	23.6424
12.5	0.5979	0.6266	−17.3472	6.8003
13.0	0.6292	0.5670	5.3295	−28.4227
13.5	0.6000	0.6219	−15.8024	4.0436
14.0	0.6436	0.6065	15.7982	−5.0842
15.0	0.6703	0.6319	35.1558	9.9431
15.5	0.6843	0.5937	45.2891	−12.6564
16.0	0.6746	0.5843	38.2966	−18.2156
17.0	0.6632	0.6153	30.0081	0.1412
18.0	0.6589	0.6236	26.8770	5.0546
19.0	0.6452	0.6316	16.9420	9.7754

$(F, \mathbf{y})$ and a given model. The final AD score is obtained by averaging the 50 runs.

Our findings can be summarized as follows. Regardless of linear or nonlinear regression, the Bayesian cut performs well with low AD scores. Introducing the unbalance and scalability factors does not deteriorate the performance of the Bayesian cut significantly. The Bayesian cut scales well when the number of change points increases.

3.2. Real Data. In this part, we apply the Bayesian cut fitting to some real data from our database shown in Table 4. This table describes feature values with regard to the increasing skeletal age that ranges from newborn to 19-year-old boys (shown in column 1) labeled by radiology experts. In order to obtain features independent of the size and the length of digits, two ratio features are used according to the paper [5]. One is L_1/L_2, the ratio of the length of distal phalanx L_1 to that of middle phalanx L_2 of the middle digit, and the other is L_2/L_3, the ratio of the length of middle phalanx L_2 to that of proximal phalanx L_3. See Figure 3 for illustration of L_1, L_2, and L_3. These two features correspond to columns 2 and 3 which are generated in the light of the algorithm in [6]. Columns 4 and 5 represent normalized values of L_1/L_2 and L_2/L_3, respectively. This normalization is done according to $(x − \mu)/\sigma$, where μ is the expectation of x and σ is the variance. In our experiments, only normalized values are used. Figure 4 shows some of the Bayesian cut fitting, where features $n(L_1/L_2)$ and $n(L_2/L_3)$ are used, models describing the relationship between the feature and the skeletal age are

the jth phase $C(ph_j)$ is randomly selected from the set $\{(C(\mathbf{f_i}) + 1), \ldots, (C(\mathbf{f_i}) + 1) + s\}$, where s is predetermined. f_i takes the value of i for $i = 1, 2, \ldots, t_k$. Note that we use a variable bound for σ_j^2 for taking into account the influence of the highest degree of the polynomial. Also, we use the variable number of sample points for each phase by introducing unbalance and scalability factors such that the performance evaluation will be more objective. To present a quantity on the performance of the Bayesian cut, we use the metric *absolute deviation* (AD), defined as

$$\text{AD} = \frac{\sum_j \left| t_j^* - t_j \right|}{k - 1}, \quad j = 1, \ldots, k - 1, \quad (4)$$

where t_j^* represents the jth element of $\mathbf{t}^*$ (the Bayesian cut). Intuitively, the smaller AD is, the closer is the Bayesian cut $\mathbf{t}^*$ to the true change points $\mathbf{t}$.

Table 3 shows the AD values. They are obtained by ranging k from 2 to 4 and s from 1 to 10. For given k, s, and a given model, 50 trials are performed to generate data, leading to 50 datasets $\{(F, \mathbf{y})\}$. We find the Bayesian cut $\mathbf{t}^*$ for each
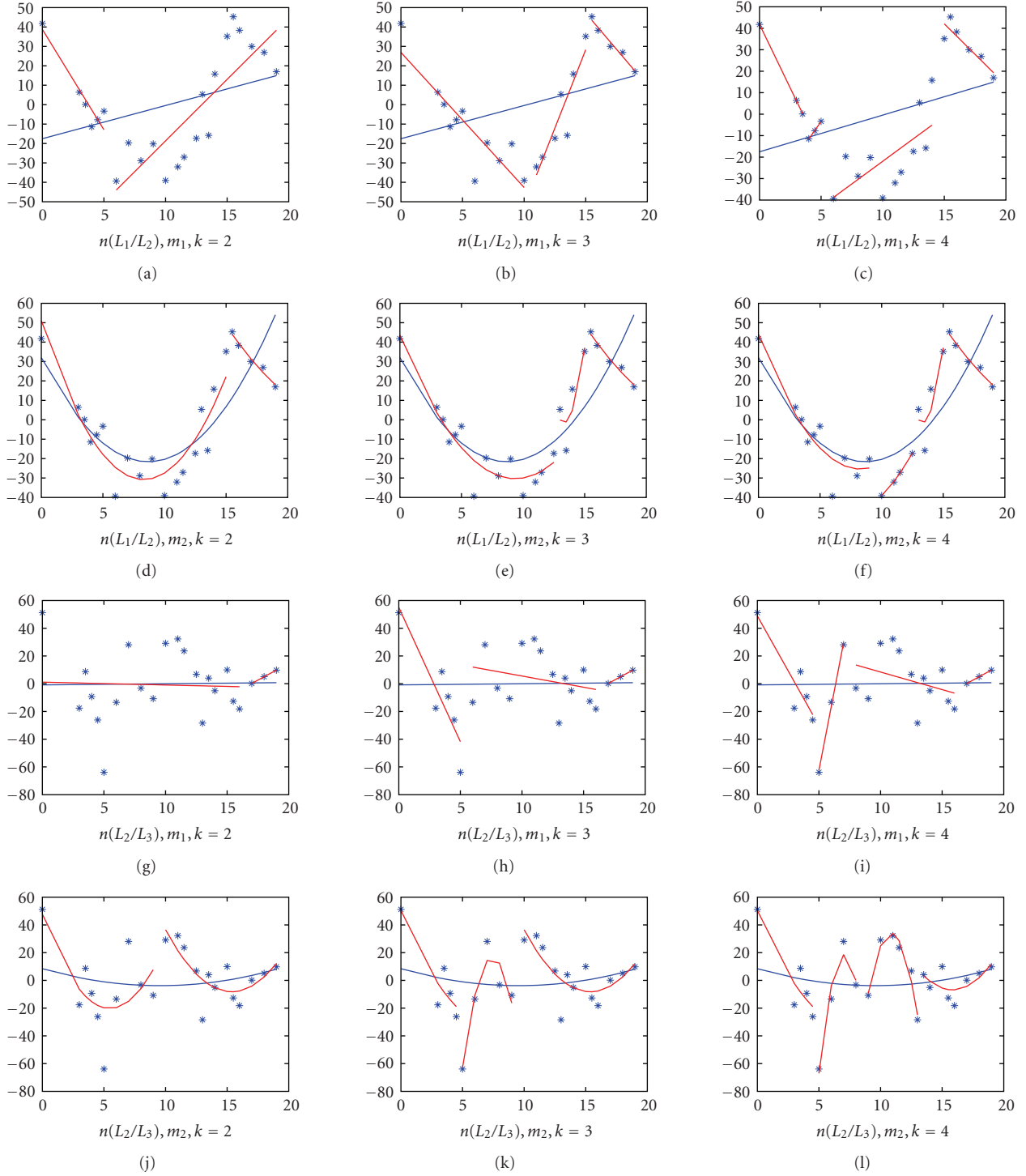
FIGURE 4: Illustration of the Bayesian cut fitting applied to the real data on features of the skeletal age.

m_1 and m_2 from Table 1, and k takes values of 2, 3, and 4. In Figure 4, the horizontal axis represents the age and the horizontal axis indicates the feature. For model m_1, the blue straight line across the entire age range is from the single (line) fitting. For model m_2, the blue curve across the entire age range is from the single (quadratic) fitting. All red (broken) lines are from the Bayesian cut fitting.

4. Conlcusion

In this paper, we propose the Bayesian cut fitting to describe features in response to the skeletal age. In the semantic space derived by our approach, the axis of skeletal age is divided into meaningful stages, within each of which the variation pattern of a feature is consistent so that a traditional

regression technique can apply to model the relationship between the skeletal age and the feature. Our approach is inspired by the observation that the variation pattern of a feature can differ in different periods of the skeletal age. A critical issue is to determine the times or change points when the variation pattern of a feature changes. This is handled by the Bayesian cut proposed in this paper. Simulations have been used to demonstrate the efficiency of the Bayesian cut fitting in terms of recovery of change points. The experiments on real data show that given a type of relationship (e.g., linear or quadratic) between the skeletal age and a feature, the Bayesian cut fitting surpasses the traditional single fitting when the consistency of the variation pattern (over the entire skeletal age range) of the feature is suspected. One major issue which is not addressed in this paper is the determination of k, the number of stages. Selection of k depends on the given data and the practical need. We leave this as our future research work.

Appendix

A. Derivation of (3)

Proof. According to the Pythagorean theorem, we have the following likelihood

$$
l\left(\vec{\beta}_j, \sigma_j^2 \mid \mathbf{y}\right) \propto \frac{1}{\left(\sigma_j^2\right)^{(t_j - t_{j-1})/2}}
$$
$$
\times \exp\left\{-\frac{1}{2\sigma_j^2}\left[\left(y_{t_{j-1}+1} - \vec{\beta}_j^T \mathbf{f}_{t_{j-1}+1}\right)^2 \right.\right.
$$
$$
\left.\left. + \cdots + \left(y_{t_j} - \vec{\beta}_j^T \mathbf{f}_{t_j}\right)^2\right]\right\}
$$
$$
\propto \frac{1}{\left(\sigma_j^2\right)^{(t_j - t_{j-1})/2}}
$$
$$
\times \exp\left\{-\frac{1}{2\sigma_j^2}\left[S_j + \left(\vec{\beta}_j - \hat{\vec{\beta}}_j\right)^T\right.\right.
$$
$$
\left.\left. \times F_j^T F_j\left(\vec{\beta}_j - \hat{\vec{\beta}}_j\right)\right]\right\},
$$
(A.1)

where $S_j = (\mathbf{y_j} - F_j\hat{\vec{\beta}}_j)^T(\mathbf{y_j} - F_j\hat{\vec{\beta}}_j)$ and $\hat{\vec{\beta}}_j = (F_j^T F_j)^{-1}F_j^T \mathbf{y_j}$. Since ϵ_{ji} are independent of each other, the likelihood function of $\vec{\beta}_1, \ldots, \vec{\beta}_k, \sigma_1^2, \ldots, \sigma_k^2, \mathbf{t}$ is then

$$
l\left(\vec{\beta}_1, \ldots, \vec{\beta}_k, \sigma_1^2, \ldots, \sigma_k^2, \mathbf{t} \mid \mathbf{y}\right)
$$
$$
\propto \prod_j \frac{1}{\left(\sigma_j^2\right)^{(t_j - t_{j-1})/2}}
$$
$$
\times \exp\left\{-\frac{1}{2\sigma_j^2}\left[S_j + \left(\vec{\beta}_j - \hat{\vec{\beta}}_j\right)^T F_j^T F_j\left(\vec{\beta}_j - \hat{\vec{\beta}}_j\right)\right]\right\}.
$$
(A.2)

Due to the assumption of the uniform prior for $\vec{\beta}_j$, $\ln(\sigma_j^2)$ and $\mathbf{t}$, we have

$$
p\left(\vec{\beta}_1, \ldots, \vec{\beta}_k, \sigma_1^2, \ldots, \sigma_k^2, \mathbf{t}\right) \propto \frac{1}{\sigma_1^2 \cdots \sigma_k^2}.
$$
(A.3)

Using (A.2) and (A.3), we have

$$
p(\mathbf{t} \mid \mathbf{y})
$$
$$
\propto \int_{\vec{\beta}_1} \cdots \int_{\vec{\beta}_k} \int_{\sigma_1^2} \cdots \int_{\sigma_k^2} \prod_j \frac{1}{\left(\sigma_j^2\right)^{(t_j - t_{j-1})/2+1}}
$$
$$
\times \exp\left\{-\frac{1}{2\sigma_j^2}\left[S_j + \left(\vec{\beta}_j - \hat{\vec{\beta}}_j\right)^T \right.\right.
$$
$$
\left.\left. \times F_j^T F_j\left(\vec{\beta}_j - \hat{\vec{\beta}}_j\right)\right]\right\}
$$
$$
\times d\vec{\beta}_1 \cdots d\vec{\beta}_k d\sigma_1^2 \cdots d\sigma_k^2.
$$
(A.4)

Note that

$$
\int_{\vec{\beta}_j} \frac{1}{\left(\sigma_j^2\right)^{(t_j - t_{j-1})/2+1}}
$$
$$
\times \exp\left\{-\frac{1}{2\sigma_j^2}\left[S_j + \left(\vec{\beta}_j - \hat{\vec{\beta}}_j\right)^T F_j^T F_j\left(\vec{\beta}_j - \hat{\vec{\beta}}_j\right)\right]\right\} d\vec{\beta}_j
$$
$$
= \frac{\exp\left(-S_j/2\sigma_j^2\right)}{\left(\sigma_j^2\right)^{(t_j - t_{j-1})/2+1}} (2\pi)^{p/2} \left(2\sigma_j^2\right)^{p/2} \left|F_j^T F_j\right|^{-1/2}.
$$
(A.5)

This equation exploits the fact

$$
\int \exp\left\{(\mathbf{x} - \vec{\mu})^T \Sigma^{-1}(\mathbf{x} - \vec{\mu})\right\} d\mathbf{x} = (2\pi)^{p/2}|\Sigma|^{1/2}, \quad (A.6)
$$

from the normal density for the p-dimensional random vector X

$$
f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{(\mathbf{x} - \vec{\mu})^T \Sigma^{-1}(\mathbf{x} - \vec{\mu})\right\} d\mathbf{x}, \quad (A.7)
$$

where $\vec{\mu}$ is the expected value of X and Σ is the variance-covariance matrix of X.

Substituting (A.5) into (A.4), we have

$$
p(\mathbf{t} \mid \mathbf{y}) \propto \prod_j \left|F_j^T F_j\right|^{-1/2} \int_{\sigma_j^2} \frac{\exp\left(-S_j/2\sigma_j^2\right)}{\left(\sigma_j^2\right)^{(t_j - t_{j-1} - p)/2+1}} d\sigma_j^2. \quad (A.8)
$$

In addition, we have

$$
\int \exp\left(-\frac{a}{2x}\right) x^{-m/2-1} dx = 2^{m/2}\Gamma\left(\frac{m}{2}\right) a^{-m/2}, \quad (A.9)
$$

from the probability density function of $X = aU$

$$f(x) = 2^{-m/2} \left[\Gamma\left(\frac{m}{2}\right) \right]^{-1} a^{m/2} x^{-m/2-1} \exp\left(-\frac{a}{2x}\right), \quad \text{(A.10)}$$

where the constant $a > 0$ and $U^{-1} \sim \chi_m^2$.

By applying (A.9) to (A.8), we get

$$p(\mathbf{t} \mid \mathbf{y}) \propto J 2^{(n-kp)/2} \prod_j \left| F_j^T F_j \right|^{-1/2}$$

$$\times \Gamma\left(\frac{t_j - t_{j-1} - p}{2}\right) S_j^{-(t_j - t_{j-1} - p)/2}, \quad \text{(A.11)}$$

where $J = \left(\sum_t 2^{(n-kp)/2} \prod_j |F_j^T F_j|^{-1/2} \Gamma((t_j - t_{j-1} - p)/2) \times S_j^{-(t_j - t_{j-1} - p)/2} \right)^{-1}$. This completes the proof. $\square$

Acknowledgments

References

[1] D. B. Darling, *Radiography of Infants and Children*, chapter 6, Charles C. Thomas, Springfield, Ill, USA, 1st edition, 1979.

[2] A. K. Poznanski, S. M. Garn, J. M. Nagy, and J. C. Gall Jr., "Metacarpophalangeal pattern profiles in the evaluation of skeletal malformations," *Radiology*, vol. 104, no. 1, pp. 1–11, 1972.

[3] D. R. Kirks, *Practical Pediatric Imaging: Diagnostic Radiology of Infants and Children*, chapter 6, Little, Brown, Boston, Mass, USA, 1st edition, 1984.

[4] J. Kosowicz, "The roentgen appearance of the hand and wrist in gonadal dysgenesis," *The American Journal of Roentgenology, Radium Therapy and Nuclear Medicine*, vol. 93, pp. 354–361, 1965.

[5] E. Pietka, M. F. McNitt-Gray, M. L. Kuo, and H. K. Huang, "Computer-assisted phalangeal analysis in skeletal age assessment," *IEEE Transactions on Medical Imaging*, vol. 10, no. 4, pp. 616–620, 1991.

[6] E. Pietka, A. Gertych, S. Pospiech, F. Cao, H. K. Huang, and V. Gilsanz, "Computer-assisted bone age assessment: image preprocessing and epiphyseal/metaphyseal ROI extraction," *IEEE Transactions on Medical Imaging*, vol. 20, no. 8, pp. 715–729, 2001.

[7] D. Chen, M. Fries, and J. M. Lyon, "A statistical method of detecting bioremediation," *Journal of Data Science*, vol. 1, no. 1, pp. 27–41, 2003.

[8] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, New York, NY, USA, 1992.