Renewing Iterative Self-Labeling Domain Adaptation With Application to the Spine Motion Prediction

Gecheng Chen, Yu Zhou, Xudong Zhang, and Rui Tuo

Abstract—The area of transfer learning comprises supervised machine learning methods that cope with the issue when the training and testing data have different input feature spaces or distributions. In this work, we propose a novel transfer learning algorithm called Renewing Iterative Self-labeling Domain Adaptation (Re-ISDA). This work is motivated by a cervical spine motion prediction problem, with the goal of predicting the cervical spine motion of different subjects using the measurements of “exterior features”. The joint distribution of the exterior features for each subject may vary with one’s distinct BMI, sex and other characteristics; the sample size of this problem is limited due to the high experimental cost. The proposed method is well suited for transfer learning problems with limited training samples. The learning problem is formulated as a dynamic programming model and the latter is then solved by an efficient greedy algorithm. Numerical studies show that the proposed method outperforms prevailing transfer learning methods. The proposed method also achieves high prediction accuracy for the cervical spine motion problem.

Note to Practitioners—This paper is motivated by the problem of using human body exterior features (e.g., contours, surface markers) acquired by computer vision or other measurement means to determine the underlying physics or mechanics. The goal is to build a subject-specific model to predict cervical spine motion based on surface-based measurements. Conventional supervised learning and transfer learning methods fail to give an accurate prediction due to the subject-to-subject variation and/or the small sample size of the training data. This paper proposes a novel transfer learning method called renewing iterative self-labeling domain adaptation to modify the possible error of the pseudo labels and thus make the training process more stable. The proposed method improves the accuracy by over 35% on the traditional transfer learning methods. This method is applicable to limited-sample problems with different training and testing input distributions, and is particularly suitable for time-series data.

Index Terms—Transfer learning, iterative learning, dynamic programming, cervical spine.

I. INTRODUCTION

A PRIMARY task of supervised learning is to learn the input-output relationship within a training set. This learned relationship is described by a model, which can then be used to make predictions for another set of interest, e.g., the testing set. Traditional supervised machine learning algorithms learn models based only on the training data. In practice, the marginal distribution of the training inputs often differs from that of the testing set. See [1] for an example in web-document classification. When this difference is large, traditional supervised machine learning algorithms may not give accurate predictions on the testing set, even if their models fitting on the training set are good. In such a circumstance, one needs to consider a transfer learning method, which builds machine learning models based on not only the training data but also the inputs of the testing set. This work is motivated by an application in studies of human motion and biomechanics. An important topic in this area is to use human body exterior features (e.g., contours, surface markers) acquired by computer vision or other measurement means to determine the underlying physics or mechanics. Studies often focus on one body region of interest in a particular type of physical act (e.g., static vs dynamic, high-speed vs low-speed, sports vs vocational). The current study is on the cervical spine motion, a critical piece of information for understanding as well as prevention of neck pain and injury. There are two fundamentally different approaches to the measurement of cervical spine motion: 1) surface-based approach, which is most widely used and more practical but is subject to significant error sources such as soft tissue motion artifact in estimating the underlying skeletal (i.e., rigid-body) movement; 2) dynamic radiography, which provides the ‘gold standard’ measurement of skeletal movement but is costly, not well accessible and predisposes human subjects to radiation exposure. In view of the practical challenges in acquiring the dynamic radiography data, we hope to build a subject-specific machine learning model that can predict subjects’ skeletal motion (output) based only on their surface-based measurements.
measurement (input). In this supervised learning problem, we have the labeled data of only six subjects due to time and budget constraints. We choose five of the subjects as the training set to build the model and the remaining as the testing set. The joint distribution of the exterior features for each subject may vary with one’s BMI, sex and other characteristics. Therefore, we must take the difference into account when constructing surrogate models. In other words, we need to address two major challenges:

1) The constructed surrogate models should perform well in spite of the difference between joint distributions of inputs of the training and testing sets;
2) The size of the labeled samples is limited.

A supervised learning approach accounting for and coping with the discrepancy between the distributions of the training and the testing inputs, e.g., the distinct distribution of each subject’s exterior features, is usually called a domain adaptation (transductive transfer learning) method [1], [2]. In domain adaptation, we call the training set the source domain and the testing set the target domain. Although several methods are available in the domain adaptation literature, they cannot fully address the stated challenges. These existing methods mainly fall into two general categories. The first category, called the instance-based transfer learning [1], proceeds by re-sampling the training data according to suitably chosen weights to approximate the distribution of testing samples. These methods are not suitable for the problems with limited labeled samples, such as the spine motion prediction problem, because the information loss due to the undersampling is more severe in small-data problems. The second category, called the feature-based transfer learning [1], proceeds by finding transformations between the inputs of the training and testing sets so that their distributions become closer. However, how to decide a suitable transformation can be difficult and domain-dependent [1]. In addition, we find that the typical methods of this category do not perform well in the spine motion prediction problem; see Section IV.

In this article, we adopt the general idea of the iterative self-labeling domain adaptation (ISDA) method [3] and propose a novel method, called the renewing self-labeling domain adaptation (Re-ISDA), to address the aforementioned challenges. The ISDA method is a domain adaptation method for classification problems. In this work, we propose a new framework, under which a version of the ISDA method is applicable for regression problems, by formulating the regression problem as a dynamic programming model with uniformly stable algorithms [4]. We also provide a greedy algorithm to solve this dynamic programming model efficiently, which proceeds by renewing all labels of testing samples labeled in former steps. This renewing step helps avoid a potential issue of the ISDA that the possible mis-labeled samples by a weak classifier in the initial stage of the iterative learning can cause serious harm to the subsequent learning process, which is also called mis-labeling issue [5], [6].

In our synthetic numerical examples, the proposed method outperforms three prevailing domain adaptation approaches. In the spine motion prediction problem, we find that the performance of Re-ISDA is superior to the three prevailing domain adaptation approaches. These numerical results suggest that the proposed Re-ISDA method is highly promising for similar human skeletal motion prediction problems.

The remainder of this paper is organized as follows. We first review the main existing methods of domain adaptation in Section II. Then we establish a new theoretical foundation for ISDA and propose the Re-ISDA in Section III. In Section IV we study the cervical spine motion prediction problem. A numerical example is investigated in Section V. Concluding remarks and discussion are given in Section VI.

II. BACKGROUND

In this section we will introduce the problem formulation and a brief review for domain adaptation. Also we will explain the relation between domain adaptation and our motivating study.

A. Problem Formulation

We start with a general formulation of domain adaptation, as discussed in [1]. A domain is defined as $D = \{X, P(X)\}$, where $X$ stands for a feature space and $P(X)$ denotes the marginal distribution of these features. Besides, a task $T$ based on a specific domain $D$ is defined as $T = \{Y, P(Y|X)\}$ where $Y$ is the label space and $P(Y|X)$ denotes the conditional distribution. Then a domain adaptation task can be described as [1] and [2]: Given a source domain (training set) $D_s$ and learning task $T_s$, based on $D_s$, domain adaptation aims to help improve the learning of the target predictive function on target domain (testing set) $D_t$ where $X_s = X_t, P_s(X) \neq P_t(X)$ and $T_s = T_t$.

In this work, we focus on regression problems in domain adaptation. Assume that the source domain $D_s = \{(x_1^s, y_1^s), (x_2^s, y_2^s), \ldots, (x_l^s, y_l^s)\}$ and the target domain as $D_t = \{x_1^t, x_2^t, \ldots, x_{l-1}^t\}$. Besides, choose a labeled sample as a calibration point $(x_r, y_r)$. In Section III-B we will discuss how to choose this calibration point in detail. Let $(X_s, Y_s)$ and $(X_t, Y_t)$ be the joint distributions of inputs and outputs on the source and the target domains. We suppose that $P(X_s) \neq P(X_t)$, i.e., the two data sets have different input distributions. We also assume that the relationship between the input and output, characterized by the conditional distributions, are the same, i.e., $P(Y_t|X_s) = P(Y_t|X_t)$. Let $(x_r, y_r)$ and $(x_t, y_t)$ be data from the source and the target distributions, respectively. Suppose we are given a loss function $L(h; x, y)$, such as the quadratic loss $L(h; x, y) = (y - h(x))^2$. The observed data are independent copies of $(x_r, y_r)$ and independent copies of $x_t$. The problem of interest is to estimate

$$h_T := \arg\min_h \mathbb{E}L(h; x_t, y_t).$$

B. Existing Domain Adaptation Methods

We review three kinds of domain adaptation methods in this section: instance based transfer learning, feature based transfer learning and iterative self-labeling domain adaptation.
1) Instanced Based Methods: Instanced based methods, such as kernel mean matching KMM [7], aim to reduce the distance between the marginal distributions of source and target domain through re-weighting the source samples based on the following formula

\[
R_i^l(f) = E_{(x,y)\in D_s} \frac{P(X_s, Y_s)}{P(X_t, Y_t)} l(f(x), y),
\]

where \(R_i^l(f)\) is the empirical loss of predictive function \(f(\cdot)\) on \(D_s\), \(l(\cdot, \cdot)\) is a loss function, \(P(X_s, Y_s)\) and \(P(X_t, Y_t)\) are the joint distributions of the source and target domains, respectively. In other words, if we can estimate the weight \(P(X_s, Y_s)\) and re-weight the source samples, we can obtain the empirical loss of \(f(\cdot)\) on target domain. However, these methods are not suitable for the problems with limited labeled samples, such as the spine motion prediction problem, as this re-weighting does not fully utilize the information of the source samples. We refer to [8], [9], [10], [11], [12], [13], and [14] for related discussions.

2) Feature Based Methods: Feature based methods proceed by finding the optimal transformation between the inputs of the training and testing set to reduce the distance between the marginal distributions of the inputs of training and testing set. An prevailing method, transfer component analysis (TCA) [15], chooses a kernel function \(\phi(\cdot)\) to map the input of the source and target domains into the reproducing kernel Hilbert space generated by \(\phi(\cdot)\), denoted by \(\mathcal{H}\), and then minimize the distance between these two marginal distributions, i.e.,

\[
\min_{\phi} \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \phi(x_s^i) - \frac{1}{n_t} \sum_{j=1}^{n_t} \phi(x_t^j) \right\|_{\mathcal{H}},
\]

where \(x_s^i \in D_s, x_t^j \in D_t, n_s\) and \(n_t\) are the size of source and target domain respectively. In our numerical studies, we find that TCA does not perform well in the spine motion prediction problem; see Sections IV-V. Some related methods are proposed in [8], [14], [16], [17], [18], [19], [20], and [21].

3) Iterative Self-Labeling Domain Adaptation: [3] propose the ISDA based on the domain adaptation SVM (DASVM) [22] and gives a theoretical foundation for this domain adaptation classification algorithm. The procedure of ISDA is illustrated in Fig. 1: assume we have some blue points as the training set (source domain) and our mission is to predict the labels of red points (target domain) as shown in the first plot of Fig. 1. In order to predict the labels for target samples more precisely, ISDA mainly repeats these three steps in its iteration process:

- First, learn a model \(f_n\) from the current labeled and pseudo labeled sample set \(S_n\). As shown in the first plot of Fig. 1, it learns a classifier through the labeled samples (blue points);
- Second, pseudo-label some target samples based on \(f_n\). If we assume it pseudo-labels one target sample each step, one of the target samples will be labeled as \(Y_1^t\) as shown in the second plot of Fig. 1;
- Then incorporate the newly labeled target sample(s) into the source samples to build up \(S_{n+1}\) to progressively modify the current classifier. As shown in the third plot of Fig. 1, the ISDA will train a new classifier based on the source samples and the newly labeled target sample, and then label a new target sample using the new classifier as \(Y_2^t\).

Note that in original ISDA, \(k\) source samples will be eliminated from the source domain at each step based on their distance from the classification boundary. In this paper we assume \(k = 0\).

Habrad proves that the final classifier will work better for domain adaptation problem if all the classifiers during the iteration work better than random guessing and at least one of these classifier is good enough. However, the possible mis-labeled samples by a weak classifier in the initial stage of the iterative learning can cause serious harm to the subsequent learning process, which is also called mis-labeling issue [5], [6].

C. Domain Adaptation in the Spine Motion Prediction Problem

In the spine motion prediction problem, the joint distribution of the surface-based measurement (input features) for each subject may vary with one’s distinct BMI, sex and other characteristics (i.e., \(P_s(X) \neq P_t(X)\)). At the same time, we assume the relationship between the surface-based measurement and the skeletal movement for all subjects, characterized by the conditional distributions, are the same [23], [24] (i.e., \(P(Y_s|X_s) = P(Y_t|X_t)\)). Based on the above two assumptions, we can formulate the spine motion prediction problem as a domain adaptation problem.

We try the TKA, KMM and ISDA for our spine motion prediction problem and find that although ISDA gets lower prediction error than TCA and KMM, the precision of ISDA remains unsatisfactory, see Section IV. In view of this, we propose a new algorithm to enhance the prediction performance of ISDA.

III. RENEWING ITERATIVE SELF-LABELING DOMAIN ADAPTATION

The proposed Renewing Iterative Self-labeling Domain Adaptation (Re-ISDA) method also follows the general steps of ISDA but it renews all the pseudo labels of target samples at each step to modify their possible error. The proposed Re-ISDA proceeds by iteratively repeating the following three steps:

1) Learn a model \(f_n\) from the current labeled and pseudo labeled sample set \(S_n\);
2) Pseudo-label some new target samples and update all the pseudo labels given in previous steps based on \(f_n\);
3) Combine the newly labeled target sample(s) and the target samples with updated labels with the source samples to build up \(S_{n+1}\).
The main steps of Re-ISDA are illustrated in Fig. 2. First, we learn a predictor through the source samples (blue points). In Fig. 2, we pseudo-label only one target sample per step. Then the Re-ISDA labels one of the target samples as $Y^1$ which is shown in the second plot of Fig. 2. Next, as shown in the third plot of Fig. 2, the ISDA will train a new predictor using the source samples and the newly labeled target sample with its label $Y^1$. Based on this new predictor, it will label a new target sample using the new predictor as $Y^2$ and update the pseudo-label $Y^1$ by $Y^2$ which means the label for the first target sample in the second iteration. By adding this updating strategy, the Re-ISDA has the ability to modify the possible target sample in the second iteration. By adding this updating strategy, the Re-ISDA has the ability to modify the possible target sample in the second iteration.

A. The Proposed Methodology

The theoretical foundation of ISDA given by [3] is not applicable in the current context, because it focuses on classification problems while we are interested in regression. Before introducing our dynamic programming framework for the domain adaptation regression problems. Some implementation guidelines for the proposed method are provided in Section III-B.

Definition 1: An algorithm $A$ has a uniform stability $\beta$ with respect to the loss function $l$ if $\forall S \in \{X \times Y\}^m, \forall i \in 1, 2, \ldots, m$

$$\sup_{(s, y) \in S} |l(h_S(x), y) - l(h_{S\setminus i}(x), y)| < \beta,$$  

(3)

where hypothesis $h_S$ is learned from $S$ and $S\setminus i$ is obtained from $S$ with deleting the $i$-th sample.

It is also shown that $\beta$ decreases at the order of magnitude $O(1/m)$ if the algorithm is stable, where $m$ is the size of $S$. So we can denote it as $\beta_S$ [4]. According to this definition, we can get an equivalent proposition:

Proposition 1: Assume that there exists a function $f$ with uniform stability $\beta$ which can fit a big enough labeled set $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, then for each point $(x, y) \in S$,

$$\sum_{f \in F} |f(x) - y| < n\beta_0,$$  

(4)

where $F = \{f = T(s) | s \subseteq S \setminus (x, y)\}$, $f = T(s)$ means $f$ is trained based on the set $s$, the size of every $s$ is big enough, $n$ is the number of such $f$, $\beta_0 = \max\{\beta_s | s \subseteq S \setminus (x, y)\}$.

We only consider the big enough subsets of $S$ in Proposition 1 because when the size of $s$ is too small, the corresponding $\beta$ could be arbitrarily large, which makes the left hand side of (4) intractable. If we define the left hand of (4) as the point loss of a fixed point in $S$, Proposition 1 can be interpreted as: if a big enough labeled set can be fitted by a uniformly stable function, the point loss is small for each fixed point in this set.

Now consider a big enough set $S = D_s \cup D_t$, where $D_s$ denotes the labeled source samples and $D_t$ denotes the target samples $D_t$, together with their true labels $Y_{true}$. Based on the assumption of domain adaptation, all samples in $S$ are generated by a single function $f$. Here we assume that the function $f$ is uniformly stable, and consequently, for each point in $S$, the corresponding point loss should be small.

Returning to our objective, we lack knowledge of the true labels $Y_{true}$, and we want to accurately predict labels $Y_i$ for the target domain $D_t$. Let $D_{tl}$ be the labeled target domain based on $Y_i$. An equivalent problem is to find labels $Y_t$ such that $D_{tl} \cup D_s$ can be fitted by a uniformly stable function, i.e., find labels $Y_i$ to minimize the point loss of a fixed point $(x_c, y_c)$, that is,

$$\min_{Y_t} \sum_{f \in F} |f(x_c) - y_c|,$$  

(5)

where $F = \{f = T(s) | s \subseteq D_s \cup D_t\}$, $D_{tl}$ is the labeled target domain by $Y_t$, the size of $s$ is big enough. We refer to point $(x_c, y_c)$ as the calibration point. Note that the calibration point is a training sample with its true label and we will discuss how to choose this point in Section III-B.

One may note that the size of $F$ in (5) may be too large, which makes this problem intractable. To address this challenge, we propose an approximation scheme. We define a proper subset of $F$, denoted as $\tilde{F}$. Then instead of solving (5), we consider the approximated problem

$$\min_{Y_t} \sum_{f \in \tilde{F}} |f(x_c) - y_c|,$$  

(6)

Next, we will model the problem (6) as a dynamic programming problem and illustrate how to define $\tilde{F}$. Let’s first fix the order of the target samples. Define state $s_\eta = \{u_1^\eta, u_2^\eta\}$, where $u_1$ is the set of labeled or pseudo-labeled samples, and $u_2$ is the set of $\eta$ newly unlabeled target samples in each state. Also we define action function as $a_n, y_c^\eta$ determined by $a_n$ is the set of pseudo labels for $u_2^\eta$. The loss at each state caused by the action $a_n$ is denoted as

$$r(s_\eta, a_n) = |f_n(x_c) - y_c|,$$  

(7)

where $f_n = T(u_1^\eta \cup (u_2^\eta, y_c^\eta))$. Figure 3 shows the first two states of the whole system, where $X_\eta = \{x_0, x_{\eta+1}, \ldots, x_{n(\eta+1)-1}\}$, $n = 0, 1, \ldots, p/\eta - 1$. For example, at state $S_0$, $u_1^0$ represents $D_s$, $u_2^0$ stands for $X_0$, pseudo labels $y_2^0$ for $X_0$ is determined by action $a_0$ and the loss $r(s_0, a_0) = |f_0(x_c) - y_c|$, where $f_0 = T(D_s \cup (X_0, y_2^0))$.  

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Our goal is to find the optimal policy \( \{a_n\}, n = 0, 1, \ldots, p/\eta - 1 \), that minimizes the sum of loss, i.e.,

\[
\min_{a_n} \sum_{i=0}^{n} r(s_n, a_n).
\tag{8}
\]

Now we consider problem (6) on the set \( F := \{f_0, f_1, \ldots, f_n\}, n = 0, 1, \ldots, p/\eta - 1 \). Then the problem is equivalent to the dynamic programming problem (8).

Solving (8) exactly is usually still computationally intractable. Thus it is reasonable to consider approximate algorithms to tackle (8). The ISDA can be regarded as an simple greedy algorithm to cope with (8), that is, to solve

\[
\min_{a_n} r(s_n, a_n),
\tag{9}
\]

step by step, for \( n = 0, 1, \ldots, p/\eta - 1 \) if we assume the action function as follows

\[
y_n^a = f_n^a(u_n^a),
\tag{10}
\]

where \( f_n^a = T(u_n^a \cup \{(x_c, y_c)\}) \) and \( T \) is a stable basic learner. This function means if we need to find suitable labels for \( u_n^a \) to minimize the loss in (7) we can just use the predictor trained by \( (u_n^a \cup \{(x_c, y_c)\}) \) to predict the labels. And the solution to the previous step is assumed as fixed and moved in the subsequent steps. The detailed process can be seen as Figure 4. At the state \( s_0 \), the pseudo label for \( X_0 \) is given by the action \( a_0 \) where the predictor \( f_0^a \) is trained from \( D_0 \cup \{(x_c, y_c)\} \), i.e., \( f_0^a = T(u_0^a \cup \{(x_c, y_c)\}) \). At the same time, the action \( a_0 \) will result in a loss \( r_0 \). Then the iteration continues.

As a simple greedy algorithm, the ISDA has the following deficiency: once a pseudo labeling error occurs, it can never be corrected. If the errors at early iterations are large, the entire downstream process can be highly disturbed, which is also called mis-labeling issue [5], [6]. This phenomenon will be shown in our experiments.

We propose a modification of the ISDA action function, which enables a self-correcting capability of the method and stabilizes the iteration process. Let \( u_n^m = \{(x_t, y_t)\} \), where \( d_n = (d_n^{l-1}, y_n^{p/l-1}) \), \( d_n^{l-1} \) is the set of target samples added in former steps, and \( y_n^{p/l-1} \) is the set of their pseudo labels given by former steps. The modified action function consists of two components:

\[
y_n^a = f_n^a(u_n^a),
\tag{11}
\]

and

\[
y_n^{p/l} = f_n^{p/l}(d_n^{l-1}),
\tag{12}
\]

where \( f_n^a = T(u_n^a \cup H^a) \) and \( T \) is a stable basic learner. In other words, in each epoch, we first pseudo label the newly added target samples, and then renew the pseudo labels of target samples added before. The first step (11), following the ISDA algorithm, explores the labels for the newly added points. The second step (12) updates the former pseudo labels in order to correct any possible errors. Note that (12) will not dramatically change the pseudo labels of former points because the difference between \( y_n^{p/l} \) and \( y_n^{p/l-1} \) should be small, which promises a stable iteration process. We call this new method Renewing Iterative Self-labeling Domain Adaptation (Re-ISDA).

The iteration process of Re-ISDA is illustrated in Fig. 5. For example, at the state \( s_1 \) there will be two actions: \( a_1 \) and \( a_1^* \); \( a_1 \) gives the pseudo label of \( x_1 \) based on the predictor \( f_1^a \) trained on \( \{x_1, y_1\} \); \( a_1^* \) renews the label \( y_1^a \). By \( f_1^{p/l} \) based on \( f_1^a \).

**Algorithm 1 Training Steps for Re-ISDA**

**Input:** ordered training set \( S = \{((x_1, y_1), (x_2, y_2), \ldots, (x_\eta, y_\eta))\} \), a labeled calibration point \((x_c, y_c)\), testing samples \( X_t = \{x_1, x_2, \ldots, x_\xi\} \), the number of newly added testing samples \( \eta \), number of epochs \( P \), initial iterative set \( S_0 = S \cup \{(x_c, y_c)\} \), a basic learner \( Q \).

**Output:** Final predictor \( f(p)(\cdot) \)

for \( p \) in 1 : \( P 

Train a predictor \( \hat{f}(p)(\cdot) \) based on \( S_{p-1} \) and \( Q \)

Pseudo-predict set \( G_p = \{((x_1, \hat{f}(p)(x_1)), (x_2, \hat{f}(p)(x_2)), \ldots, (x_\xi, \hat{f}(p)(x_\xi))\} \)

Renew iterative set \( S_p = S_0 \cup G_p \)

end for

The detailed algorithm of Re-ISDA is shown in Algorithm 1. We need to choose a basic learner \( Q \), the calibration point \((x_c, y_c)\), the order of adding target samples and the size of added samples \( \eta \) at each step when implementing the Re-ISDA. Some suggestions for implementation details of the proposed algorithm will be provided in Section III-B.
B. Implementation Details for Re-ISDA

In this section, we offer some practical guidelines for implementing the proposed Re-ISDA method.

1) Basic Learner Q: We recommend using feed-forward neural networks (FNNs) as a basic learner. According to [25], FNNs are robust learners. Besides, FNNs can handle huge data sets and do not require complex transformation or representation learning for the raw data before training [26]. In order to get a well-performing FNN, one needs to choose suitable hyper-parameters such as the number of layers and nodes, learning rate, the number of training epochs, and so on. Detailed tuning guidelines for FNNs can be found in [27], [28], and [29].

2) Calibration Points (x_c, y_c): In domain adaptation problems we may face two different situations: 1) no labeled target samples can be used; 2) a few labeled target samples are available. Here we give some guidelines for choosing the calibration points under different occasions:

- If we only have the labels for the source domain, we recommend choosing the sample source as the calibration point which has the least average distance between the target samples;
- If we only have one labeled target sample, we can choose this one as the calibration point;
- If we have several labeled target samples, we can choose one point randomly from them as the calibration point or implement Re-ISDA for each labeled target sample and ensemble their results to give the final prediction results.

3) Order of Added Target Samples: We make the suggestions below for the order of added target samples under two types of data:

- If the data is in a time-series format, we should maintain the original order to keep the dynamic properties.
- If the data does not have a temporal structure, we recommend sorting the target samples from small to large according to their distance to the calibration point.

4) Size of Added Samples in Each Step $\eta$: The numerical experiments in Section IV show that the Re-ISDA with smaller $\eta$ will lead to more precise prediction results and a more stable iteration process. On the other hand, a smaller $\eta$ will require more iteration steps and thus increase the computational cost. Therefore, we recommend setting $\eta$ as small as possible provided that the computational cost is affordable.

IV. CERVICAL SPINE MOTION PREDICTION

In this section, we apply the proposed Re-ISDA method to the cervical spine motion prediction problem. The numerical results show that the Re-ISDA is more suitable for the cervical spine motion prediction problem compared with some prevailing transfer learning methods.

A. Background

Cervical spine, a highly complex multi-joint structure, supporting the head weight and providing the mobility and flexibility of the head [30], is susceptible to injuries that occur acutely and chronically. To design effective injury prevention and control, an accurate characterization of the cervical spine motion is needed for a better understanding of the neck biomechanics as well as the pathomechanics of neck injury and neck pain [30], [31], [32].

Conventional approaches to measuring in vivo three-dimensional (3D) skeletal motion and position often involve using optical markers systems and inertial sensors placed on the body surface. Such surface-based measurements are subject to soft tissue motion artifacts and therefore do not reflect the underlying skeletal motion accurately [33], [34], [35], [36]. In addition, marker-based or sensor-based approaches are more suited for measuring the kinematics of long bones (i.e., extremities). For structures such as the cervical spine where relatively small individual bones are inter-connected by complex joints, surface-based measurements are simply too error-prone. Dynamic stereo-radiography (DSX) has proven to be an accurate, effective, and non-invasive method of quantifying 3D in vivo cervical spine motion and is regarded as the gold standard. However, the technology is not widely accessible and incurs tremendous cost both in development and use.

Previous studies have focused on using surface markers to estimate 3D bone motion or accessing the correlations between measurements from surface markers and from rigid markers (i.e., the markers are mounted on the bones) but are all limited to the lower extremity [37], [38], [39]. In a recent study, Nichoson et. al developed machine learning algorithms to predict scapular kinematics using optical motion measurement data [40], which further inspires us to use statistical or machine learning approaches to estimate and predict 3D cervical spine motion using optical motion measurements. This and our recent success in measuring cervical spine motion using DSX and optical motion capture inspire us to pursue new machine learning approaches to estimate cervical vertebrae motion.

B. Experimental Data Acquisition

In this experiment, participants performed self-paced head-neck flexion-extension tasks (i.e., dynamic free range-of-motion tasks) with the head-neck moving primarily in the sagittal plane while their cervical spine region was imaged continuously using a DSX system [41]. Informed consent approved by the Institutional Review Board of the University of Pittsburgh (where the experiments were conducted) was obtained from all participants. A 12-camera Vicon motion capture system (Vantage-Series, Vicon Motion Labs, Oxford, UK) was used to monitor and record the head-neck motions from retro-reflective spherical surface markers placed on ten anatomical landmarks: glabellae, inferior border of each orbit (left and right), the tragion notches (left and right), acromion processes (left and right), suprasternal notch, sternum, and C7 spinous process as shown in Fig. 6.

The recorded trajectories of the surface markers were tracked in a global laboratory coordinate system, resulting in time-histories of three coordinates (x, y, and z) in each marker.

Three-dimensional (3D) orientation and position of each vertebra were determined via a previously validated volumetric model-based tracking algorithm [42], [43], [44], [45]. Cervical sagittal spinal curvature was measured at each time
C7-MMP as the input data to predict the corresponding Cobb angle. In contrast to the BMI information and C7-MMP, which do not evolve in time, the rest of the input data are time series. They capture the position of the corresponding markers and are collected sequentially with fixed time intervals.

As a pre-processing step, we take the difference of every two consecutive frames of the time-series data. We believe that this difference can be used to enhance the prediction performance because it reflects the velocity of each marker. We include this computed “velocity information” of the time-series input variables as part of the input data for each candidate learning method. In summary, the input dimension is 62.

D. Objective and Problem Formulation

We have labeled information (both marker information and its corresponding Cobb angle information) of six people: subject #3, 4, 17, 31, 34, 38 with 96, 96, 95, 89, 89, 95 labeled samples, respectively. The goal is to identify a suitable method to predict the Cobb angle information based on the corresponding marker information, which is required and necessary for a better understanding of neck biomechanics. To this end, we need to partition the dataset into two parts, i.e., the source domain and the target domain, to test the methods we have. Here we test the prediction performance for each subject by considering one subject as the target domain and all other subjects as the source domain. For example, We divide the dataset as follows if we consider the #17 subject as the target domain:

• Source domain (training set): subject #3, 4, 31, 34, 38 (called $D_{s1}, D_{s2}, D_{s3}, D_{s4}, D_{s5}$);
• Target domain (testing set): subject #17 (called $D_t$).

Additionally, we assume that a labeled target sample $\{x^*, y^*\}$ is available (working as the calibration point), which is the first sample of each subject with its true label. This assumption is reasonable in real-world applications because we only need one frame of the biplane radiography data, which can be collected via a conventional static radiography equipment in an inexpensive experiment.

In this work, we consider four candidates of domain adaptation methods: the proposed Re-ISDA and other three prevailing transfer learning methods. The performance of each method is measured by the RMSE under the testing data.

E. Data Analysis

First, we apply the max-min normalization on both the source and target domain data set to stabilize the forthcoming
training process. Next, we employ the principal component analysis (PCA) to reduce the dimensions and choose the first ten components as the new design.

Fig. 8 shows the distribution of the first two components of #3, 4, 31, 34, 38, 17 against their true Cobb angle information, which can approximately reflect their distribution in the 63 dimensions space. From Fig. 8, the data from each subject can be fitted by a separate curve. Thus, we consider the task of prediction as a multi-source transform learning problem. Denote the function trained by each source as $f_i = T(D_{si}), i = 1, 2, 3, 4, 5,$ and the prediction function for target domain $D_t$ as $f_t$. For convenience, we choose $f_t$ according to the following rule:

$$f_t = \arg \min_{f_i} | f_i(x^*) - y^* |,$$  (13)

A neural network with structure $c(10, 16, 16, 8, 1)$ is used as the basic learner for these five methods. Its learning rate is 0.1, and the number of training epochs is 1500. The implementation details of Re-ISDA are as follows:

- The basic learning of Re-ISDA is the neural network with the same structure.
- We use the first sample of the target domain with its true label as the calibration point as we discussed in Section IV-D.
- The order of added target samples just remains their original order because they have a time-series structure.
- Two new target samples will be added in each step of the training process.

The RMSEs of the five methods for different subjects are given in Table I. The pointwise predictive value for every point of subject #17 is shown in Fig. 9 where the black points represent the true values. Clearly, we can see from the Fig.9:

- The brown line (NN) deviates far away from the black one (true value), especially for the latter half, representing that the traditional predictor can not work well for this situation;
- The performance of the purple line (KMM) is worse than NN. In this experiment, the difference between weights of different source samples is up to $10^{10}$. As a result, many points with tiny weights are actually deleted from the training set, which causes the under-fitting of the model;

| Subject | NN without DA | TCA | KMM | ISDA | Re ISDA |
|---------|---------------|-----|-----|------|---------|
| # 3     | 26.52         | 27.01 | 33.72 | 26.42 | 24.32   |
| # 4     | 12.92         | 24.03 | 18.23 | 9.03  | 8.51    |
| # 17    | 8.19          | 26.34 | 13.2  | 6.21  | 2.68    |
| # 31    | 8.00          | 24.03 | 28.21 | 7.92  | 7.68    |
| # 34    | 8.18          | 16.23 | 48.02 | 6.67  | 6.30    |
| # 38    | 14.84         | 44.01 | 38.02 | 12.02 | 6.97    |

Fig. 10. Performance of Re-ISDA for different $\eta$ during their training processes.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
CHEN et al.: RENEWING ITERATIVE SELF-LABELING DOMAIN ADAPTATION 3551

Fig. 11. How the predictive labels in ISDA versus the labels in Re-ISDA change over iterations while more and more target samples are added.

• The blue line (TCA) is approximately a straight line. We have tried different kernel functions and their parameters for the TCA but get some straight lines just like the line shown here. The possible reason might be that the TCA makes the source domain harder to learn for a regression mission;
• The green line (ISDA) works much better than the above three methods and is very close to the black line. However, its performance for the former half part is not desirable. The possible reason for it might be the little flexibility of ISDA so that it can not modify the labels of the former samples based on the points coming afterward;
• The red line (Re-ISDA) is approximately the same as the black one, which indicates that our method works highly well. This result shows that the proposed Re-ISDA method can transfer the predictor through different marginal distributions and is flexible enough to modify the pseudo labels in the former steps.

F. In-Depth Analysis for the Performance of Re-ISDA

In this section, we study the performances of Re-ISDA under different scenarios to show its convergence and provide some empirical guidelines.

First, we test the performance of Re-ISDA for different $\eta$. Fig. 10 shows the RMSEs for the labeled points (both newly labeled and labeled in previous steps) at each step of Re-ISDA when $\eta = 2, 3, 5$. In the iteration processes, the RMSEs of predictors increase at the beginning and then fall, proving that the Re-ISDA modifies the prediction error of the added points through the new coming samples. Besides, if we choose the smaller $\eta$, the performance of the predictor changes more smoothly and ends at a lower RMSE. However, a smaller $\eta$ means more training epochs and higher computation cost, so that one needs to make a trade-off between precision and cost.

Second, we compare the predictive labels generated by ISDA and Re-ISDA during the training process for subject # 17, in order to demonstrate how the proposed Re-ISDA corrects the labeling errors made by ISDA. Figure 11 displays the predictive labels for different numbers of target samples. The red and green lines indicate the predictive labels of Re-ISDA and ISDA, respectively. It can be observed that during the initial stages of training (sub-figures a-d in Figure 11), the labels generated by both methods are nearly identical. However, in the latter half of the training process, the red lines gradually converge towards the black lines (representing the true labels) with the consideration of more target samples, particularly for the 10th - 50th target samples. This figure highlights that the proposed Re-ISDA effectively rectifies the possible labeling errors of ISDA and mitigates the issue of mislabeling. Furthermore, it is evident from this figure that Re-ISDA has the ability to rectify the incorrect (or insufficiently precise) paths provided in previous iterations by taking into account additional information about the target samples. Additionally, this figure also shows that the Re-ISDA is able to fix the incorrect path given by previous iterations by considering more target samples.

Next, we evaluate the performance of Re-ISDA under different training conditions, where different numbers of source samples are used. The # 17 subject is used as the target domain for this experiment. Figure 12 presents the predictive labels generated by Re-ISDA (depicted by the red lines) for different numbers of source samples used during training. It can be observed that the predictive labels are not in close agreement with the true labels (represented by the black lines) when only 24 source samples are used. However, Re-ISDA still captures the general trend of the true labels. As more source samples are incorporated during training, the performance of Re-ISDA gradually improves and finally, the red line converges to the black line.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Finally, in order to show the convergence property of Re-ISDA, we implement it ten times for the # 17 subject, starting from different initial weights of the basic learner. Figure 13 shows the results where the black line means the true labels and red lines represent the predictive labels of Re-ISDA. One can see that the predictive labels of Re-ISDA starting from different weights are centered around the true labels.

V. SIMULATION STUDIES

In the Supplementary, we include a comprehensive toy example to show the superiority of the proposed Re-ISDA.

VI. CONCLUSIONS AND DISCUSSION

This work proposes a novel method called Re-ISDA for the domain adaptation problem with limited labeled samples based on the ISDA algorithm. We first formulate the learning problem as a dynamic programming model and then propose a greedy algorithm to solve it efficiently. The proposed method updates all labels of the target samples (both newly added and added in former iterations) at each iteration, leading to higher flexibility and fewer mis-labeling issues than ISDA. Our numerical experiments and a cervical spine motion prediction example confirm the prediction performance of the proposed method.

The computational cost of the proposed method is slightly higher than the other domain adaptation methods we have compared with. Thus the proposed method should be more suitable for moderate sized data. Besides, the Re-ISDA is more suitable for the time-series data because it can incorporate the time-series information. In this work, we only consider prediction problems, but the proposed method, with proper modifications, should also be applicable to classification problems.

Given that the Re-ISDA is based on a greedy algorithm, one may doubt whether Re-ISDA is guaranteed to converge during the iterations. According to [49] and [50], many successful regression and classification algorithms can be regarded as greedy procedures for stagewise minimization of the corresponding loss function. These two papers show that the local optima is good enough if the global optima is hard, even impossible, to achieve.

Additionally, based on Definition 1 and Proposition 1, the difference between the predictive labels and the true labels in each step should be bounded by a small number. That is, the Re-ISDA will not find a path quite far away from the true path (i.e., diverge). Of course, it is possible for the Re-ISDA to move along an insufficiently accurate path, but we still show that the Re-ISDA can fix the inaccurate path during the training when more and more target samples are considered empirically in Section IV-F. We will consider combining addition approaches, i.e., probabilistic measures, to make the training process of Re-ISDA more stable and accurate in our future work.

REFERENCES

[1] S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[2] I. Redko, E. Morvant, A. Habrard, M. Sebban, and Y. Bennani, Advances in Domain Adaptation Theory. Amsterdam, The Netherlands: Elsevier, 2019.
[3] A. Habrard, J.-P. Peyrache, and M. Sebban, “Iterative self-labeling domain adaptation for linear structured image classification,” Int. J. Artif. Intell. Tools, vol. 22, no. 5, Oct. 2013, Art. no. 1360005.
[4] O. Bousquet and A. Elisseeff, “Stability and generalization,” J. Mach. Learn. Res., vol. 2, pp. 499–526, Mar. 2002.
[5] Q. Wang and T. Breckon, “Unsupervised domain adaptation via structured prediction based selective pseudo-labeling,” in Proc. AAAI, 2020, vol. 34, no. 4, pp. 6243–6250.
[6] J. Zhang, W. Li, and P. Ogunbona, “Joint geometrical and statistical alignment for visual domain adaptation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 5150–5158.
[7] J. Huang, A. Gretton, K. Borgwardt, B. Scholkopf, and A. J. Smola, “Correcting sample selection bias by unlabeled data,” in Proc. Adv. Neural Inf. Process. Syst., 2007, pp. 601–608.
[8] L. Zhang and X. Gao, “Transfer adaptation learning: A decade survey,” IEEE Trans. Neural Netw. Learn. Syst., early access, Jun. 21, 2022, doi: 10.1109/TNNLS.2022.3183326.
[9] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, “Boosting for transfer learning,” in Proc. 24th Int. Conf. Mach. Learn., Jun. 2007, pp. 193–200.
[10] D. Chattopadhayay, Q. Sun, W. Fan, I. Davidson, S. Panchanathan, and J. Ye, “Multisource domain adaptation and its application to early detection of fatigue,” ACM Trans. Knowl. Discovery Data, vol. 6, no. 4, pp. 1–26, Dec. 2012.
[11] H. Daumé, “Frustratingly easy domain adaptation,” 2009, arXiv:0907.1815.
[12] L. Duan, D. Xu, and I. W. Tsang, “Domain adaptation from multiple sources: A domain-dependent regularization approach,” IEEE Trans. Neural Netw. Learn. Syst., vol. 23, no. 3, pp. 504–518, Mar. 2012.
[13] Q. Chen, Y. Liu, Z. Wang, I. Wassell, and K. Chetty, “Re-weighted adversarial adaptation network for unsupervised domain adaptation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7976–7985.
[14] F. Zhang et al., “A comprehensive survey on transfer learning,” Proc. IEEE, vol. 109, no. 1, pp. 43–76, Jan. 2021.
[15] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain adaptation via transfer component analysis,” IEEE Trans. Neural Netw. Learn. Syst., vol. 22, no. 2, pp. 199–210, Feb. 2011.
[16] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, “Transfer feature learning with joint distribution adaptation,” in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2013, pp. 2200–2207.
[17] L. J. Pan, X. Ni, J. T. Sun, Q. Yang, and Z. Chen, “Cross-domain sentiment classification via spectral feature alignment,” in Proc. 19th Int. Conf. World Wide Web, Apr. 2010, pp. 751–760.
[18] J. Giles, K. K. Ang, L. S. Mihaylova, and M. Arvaneth, “A subject-to-subject transfer learning framework based on Jensen–Shannon divergence for improving brain-computer interface,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 3087–3091.
[19] D. Li, Y. Zhang, X. Liu, and K. Zhao, “Regularised transfer learning for hyperspectral image classification,” IET Comput. Vis., vol. 13, no. 2, pp. 188–193, Mar. 2019.
[20] L. Zhang, J. Liu, Y. Yang, F. Huang, F. Nie, and D. Zhang, “Optimal projection guided transfer hashing for image retrieval,” IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 10, pp. 3788–3802, Oct. 2020.
[21] Y. Qin, Q. Qian, Y. Wang, and J. Zhou, “Intermediate distribution alignment and its application into mechanical fault transfer diagnosis,” IEEE Trans. Ind. Informat., vol. 18, no. 10, pp. 7305–7315, Oct. 2022.
[22] L. Bruzzone and M. Marconcini, “Domain adaptation problems: A DASVM classification technique and a circular validation strategy,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 5, pp. 770–787, May 2010.

[23] C. K. Anderson, D. B. Chaffin, and G. D. Herrin, “A study of lumbar-sacral orientation under varied static loads,” Spine, vol. 11, no. 5, pp. 456–462, Jun. 1986.

[24] X. Zhang and J. Xiong, “Model-guided derivation of lumbar vertebral kinematics in vivo reveals the difference between external marker-defined and internal segmental rotations,” J. Biomechanics, vol. 36, no. 1, pp. 9–17, Jan. 2003.

[25] H. Xu and S. Mannor, “Robustness and generalization,” Mach. Learn., vol. 86, no. 3, pp. 391–423, Mar. 2012.

[26] E. Akleman, “Deep learning,” Computer, vol. 53, no. 9, p. 17, Sep. 2020.

[27] J. Schmidhuber, “Deep learning: Methods and applications,” Neural Netw., vol. 61, pp. 85–117, Jan. 2015.

[28] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016.

[29] L. Deng and D. Yu, “Deep learning: An overview,” in Trends Signal Process., vol. 7 nos. 3–4, pp. 197–387, 2013.

[30] N. Bogduk and S. Mercer, “Biomechanics of the cervical spine. I: Normal kinematics,” Clin. Biomechanics, vol. 15, no. 9, pp. 633–648, Nov. 2000.

[31] D. F. Huelke and G. S. Nusholtz, “Cervical spine biomechanics: A review of the literature,” J. Orthopaedic Res., vol. 4, no. 2, pp. 232–245, 1986.

[32] E. E. Swartz, R. Floyd, and M. Cendosa, “Cervical spine functional anatomy and the biomechanics of injury due to compressive loading,” J. Athletic Training, vol. 40, no. 3, p. 155, 2005.

[33] A. Cappozzo, F. Catani, A. Leardini, M. Benedetti, and U. D. Croce, “Position and orientation in space of bones during movement: Experimental artefacts,” Clin. Biomechanics, vol. 11, no. 2, pp. 90–100, Mar. 1996.

[34] K. Li, L. Zheng, S. Tashman, and X. Zhang, “The inaccuracy of surface-measured model-derived tibiofemoral kinematics,” J. Biomechanics, vol. 45, no. 15, pp. 2719–2723, Oct. 2012.

[35] T.-Y. Tsai, T.-W. Lu, M.-Y. Kuo, and C.-C. Lin, “Effects of soft tissue artifacts on the calculated kinematics and kinetics of the knee during stair-ascent,” J. Biomechanics, vol. 44, no. 6, pp. 1182–1188, Apr. 2011.

[36] D. L. Benoit, D. K. Ramsey, M. Lamontagne, L. Xu, P. Wretenberg, and P. Renström, “Effect of skin movement artifact on knee kinematics during gait and cutting motions measured in vivo,” Gait Posture, vol. 24, no. 2, pp. 152–164, Oct. 2006.

[37] A. Cappozzo, A. Cappello, U. D. Croce, and F. Pensalfini, “Surface-marker cluster design criteria for 3-D bone movement reconstruction,” IEEE Trans. Biomed. Eng., vol. 44, no. 12, pp. 1165–1174, Dec. 1997.

[38] A. Ceretti, U. Delia Croce, and A. Cappozzo, “Reconstruction of skeletal movement using skin markers: Comparative assessment of bone pose estimators,” J. NeuroEng. Rehabil., vol. 3, no. 1, pp. 1–12, Dec. 2006.

[39] T. Fukaya, H. Mutsuzaki, H. Ida, and Y. Wadano, “Two different protocols for knee joint motion analyses in the stance phase of gait: Correlation of the rigid marker set and the point cluster technique,” Rehabil. Res. Pract., vol. 2012, Jan. 2012, Art. no. 586348.

[40] K. F. Nicholson, R. T. Richardson, E. A. R. van Roden, R. G. Quinton, K. F. Anzilotti, and I. G. Richards, “Machine learning algorithms for predicting scapular kinematics,” Med. Eng. Phys., vol. 65, pp. 39–45, Mar. 2019.

[41] Y. Zhou et al., “A state-of-the-art integrative approach to studying neck biomechanics in vivo,” Sci. China Technol. Sci., vol. 63, no. 7, pp. 1235–1246, Jul. 2020.

[42] W. J. Anderst, E. Baillargeon, W. F. Donaldson, J. Y. Lee, and J. D. Kang, “Validation of a noninvasive technique to precisely measure in vivo three-dimensional cervical spine movement,” Spine, vol. 36, no. 6, pp. 393–400, Mar. 2011.

[43] M. J. Bey, R. Zauel, S. K. Brock, and S. Tashman, “Validation of a new model-based tracking technique for measuring three-dimensional, in vivo glenohumeral joint kinematics,” J. Biomechanical Eng., vol. 128, no. 4, pp. 604–609, Aug. 2006.

[44] W. Anderst, R. Zauel, J. Bishop, E. Demps, and S. Tashman, “Validation of a three-dimensional model-based tibio-femoral tracking during running,” Med. Eng. Phys., vol. 31, no. 1, pp. 10–16, Jan. 2009.

[45] W. J. Anderst, J. Y. Lee, W. F. Donaldson, and J. D. Kang, “Six-degrees-of-freedom cervical spine range of motion during dynamic flexion-extension after single-level anterior arthrodesis: Comparison with asymptomatic control subjects,” J. Bone Joint Surgery, vol. 95, no. 6, pp. 497–506, Mar. 2013.

[46] J. Cobb, “Outline for the study of scoliosis,” Instr Course Lect., vol. 5, pp. 261–275, Mar. 1948.

[47] D. E. Harrison, R. Cailliet, D. D. Harrison, T. J. Janik, and B. Holland, “Reliability of centroid, Cobb, and Harrison posterior tangent methods: Which to choose for analysis of thoracic kyphosis,” Spine, vol. 26, no. 11, pp. 227–234, Jun. 2001.

[48] V. Vrtovec, F. Pernus, and B. Likar, “A review of methods for quantitative evaluation of spinal curvature,” Eur. Spine J., vol. 18, no. 5, pp. 593–607, May 2009.

[49] S. Mannor, R. Meir, and T. Zhang, “Greedy algorithms for classification—Consistency, convergence rates, and adaptivity,” J. Mach. Learn. Res., vol. 4, pp. 713–742, Oct. 2003.

[50] A. R. Barron, A. Cohen, W. Dahmen, and R. A. DeVore, “Approximation and learning by greedy algorithms,” Ann. Statist., vol. 36, no. 1, pp. 64–94, Feb. 2008.

Gecheng Chen received the B.S. degree from Wuhan University, Wuhan, China, in 2014, and the M.S. degree from Zhejiang University, Hangzhou, China, in 2019. He is currently pursuing the Ph.D. degree with the Industrial and Systems Engineering Department, Texas A&M University. His research interests include machine learning and computer experiments.

Yu Zhou received the B.S. degree in mechanical engineering from Nanjing Normal University, Nanjing, China, in 2015, and the M.S. degree in mechanical engineering from the University of Pittsburgh, Pittsburgh, PA, USA, in 2017. He is currently pursuing the Ph.D. degree with the Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, USA. His research interests include studying neck biomechanics during dynamic acts and sustained exertions in vivo using an integrated approach that combines multiple imaging modalities, biodynamic measurements, and computational modeling techniques.

Xudong Zhang is currently a Professor and a Liollio Family Fellow of Industrial and Systems Engineering and a Professor of Mechanical Engineering and Biomedical Engineering with Texas A&M University. His research interests include modeling, analysis, and simulation of the human musculoskeletal system and its multi-scale structures and components during functional, particularly dynamic activities. The overarching goal is to advance the musculoskeletal biomechanical science bases and applications in healthcare and human-machine systems engineering.

Rui Tuo received the Ph.D. degree in statistics from the University of Chinese Academy of Sciences in 2013. He is currently an Assistant Professor with the Department of Industrial and Systems Engineering and an Adjunct Assistant Professor with the Department of Statistics, Texas A&M University. His research interests include data science for computer simulations, including design and analysis of computer experiments and uncertainty quantification. Recently, he has been working on theory and computational techniques for nonparametric models and Gaussian process regression.