Medical Engineering and Physics 37 (2015) 885–891

Contents lists available at ScienceDirect

Medical Engineering and Physics

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

Surrogate modeling of deformable joint contact using artificial neural networks

Ilan Eskinazi[a], Benjamin J. Fregly[a,b,c,*]

[a] Department of Mechanical & Aerospace Engineering, University of Florida, Gainesville, FL, USA
[b] Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA
[c] Department of Orthopaedics and Rehabilitation, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Article history:
Received 10 February 2015
Revised 9 June 2015
Accepted 24 June 2015

Keywords:
Elastic contact
Surrogate modeling
Neural networks
Biomechanics
Response surface
Meta-model
Knee contact
Tibiofemoral joint

ABSTRACT

Deformable joint contact models can be used to estimate loading conditions for cartilage–cartilage, implant–implant, human–orthotic, and foot–ground interactions. However, contact evaluations are often so expensive computationally that they can be prohibitive for simulations or optimizations requiring thousands or even millions of contact evaluations. To overcome this limitation, we developed a novel surrogate contact modeling method based on artificial neural networks (ANNs). The method uses special sampling techniques to gather input–output data points from an original (slow) contact model in multiple domains of input space, where each domain represents a different physical situation likely to be encountered. For each contact force and torque output by the original contact model, a multi-layer feed-forward ANN is defined, trained, and incorporated into a surrogate contact model. As an evaluation problem, we created an ANN-based surrogate contact model of an artificial tibiofemoral joint using over 75,000 evaluations of a fine-grid elastic foundation (EF) contact model. The surrogate contact model computed contact forces and torques about 1000 times faster than a less accurate coarse grid EF contact model. Furthermore, the surrogate contact model was seven times more accurate than the coarse grid EF contact model within the input domain of a walking motion. For larger input domains, the surrogate contact model showed the expected trend of increasing error with increasing domain size. In addition, the surrogate contact model was able to identify out-of-contact situations with high accuracy. Computational contact models created using our proposed ANN approach may remove an important computational bottleneck from musculoskeletal simulations or optimizations incorporating deformable joint contact models.

1. Introduction

Deformable contact models can be incorporated into multi-body dynamic simulations to compute the loads resulting from surface–surface interactions. In musculoskeletal biomechanics, the need to model deformable contact typically arises when cartilage–cartilage [1–3], implant–implant [4,5], human–orthotic [6,7], and foot–ground [8–11] interactions occur. However, deformable contact models are computationally expensive and thus can be prohibitive for studies that require large numbers of repeated contact evaluations such as optimizations and forward dynamic simulations.

Surrogate contact models can provide one solution to this problem. Surrogate models, also known as meta-models or response surface approximations, fit or interpolate input–output relationships sampled from a "slow" computational model (e.g., a finite element or elastic foundation contact model). The simplest example of a surrogate contact model is a response surface or multiple linear regression model, which has been used to calculate cartilage–cartilage contact forces in a natural tibiofemoral joint [12]. Kriging is a more complex surrogate modeling technique that has also been used to model contact forces and torques in the knee [13]. Kriging-based contact models have been used in an optimization approach that predicted muscle forces, tibiofemoral contact forces, and patellofemoral contact forces simultaneously in the knee during walking [14]. Other efforts to create surrogate knee contact models include a Hammerstein–Wiener model, a nonlinear autoregressive model with exogenous input, and a time delay artificial neural network [15]. In addition, a surrogate foot–ground contact model has been created using a lazy learning interpolation method [16].

While each of these surrogate contact modeling methods improves computational speed, each also suffers from important limitations. Kriging-based models suffer from two disadvantages. First, only a relatively low number of sample points can be interpolated given

* Corresponding author at: Department of Mechanical & Aerospace Engineering, 231 MAE-A Building, PO Box 116250, University of Florida, Gainesville, FL 32611-6250, USA. Tel.: +1 352 392 8157; fax: +1 352 392 7303.
E-mail address: fregly@ufl.edu (B.J. Fregly).

http://dx.doi.org/10.1016/j.medengphy.2015.06.006
1350-4533/

a computer's memory resources. As the number of sample points increases, so does the necessary memory and computation time required for model construction and use. Second, the most common implementation of Kriging interpolates the data instead of regressing it. This property becomes a disadvantage when the data contain noise, such as with element-based contact models whenever a proportionately small number of elements are loaded (e.g., at low loads with a coarse mesh). Hammerstein–Weinner, nonlinear autoregressive and time delay ANN models require knowledge of past configurations in time to evaluate the current configuration of the contacting bodies. This requirement implies that these methods can be used only in time-incremented analyses. Lazy learning models have the advantage of bounding the prediction error and adapting to a changing domain, but these benefits come at the cost of requiring additional "slow" contact model evaluations during surrogate model use.

To address the need for fast, accurate, and multi-purpose surrogate contact models for musculoskeletal simulations and optimizations, this study explores the use of multi-layer feed-forward ANN models. ANNs can be formulated as time-independent regression problems capable of fitting arbitrary observable functions [17]. Feed-forward ANNs have been used as black box models in previous non-contact biomechanical simulations to determine multi-dimensional input–output relationships [18,19] but have not been explored for deformable contact applications. Our proposed ANN contact modeling approach includes a special sampling technique that seeks to improve computational speed and accuracy over existing Kriging-based schemes. The method also permits fitting many more sample points than would be possible with Kriging, allowing for varying levels of accuracy across multiple domains of input space. Furthermore, the approach fits the sample points via regression rather than interpolation, effectively smoothing noise in the sampled data points [20]. The computational speed and accuracy of our ANN contact modeling approach are evaluated using an elastic foundation (EF) contact model of an artificial tibiofemoral joint.

2. Methods

2.1. Surrogate contact modeling background

The goal of surrogate contact modeling is to replace a computationally "slow" contact model with a computationally "fast" contact model that exhibits the same input–output characteristics. This process involves sampling the computationally expensive model, henceforth called the *original model*, using a sampling plan or design of experiments. Sampling yields a series of *sample points* that relate original model inputs to original model outputs. For multi-body applications, the inputs to the original model are *pose parameters* consisting of three translations and three rotations defining the position and orientation of one contacting body with respect to the other. The corresponding outputs of the original model are *contact loads* consisting of three contact forces and three torques calculated with respect to a selected point on one of the contacting bodies. Finally, testing takes place to evaluate the discrepancy between the original model and the resulting surrogate model using sample points not included in the surrogate model construction process.

The method presented in this paper builds upon a previous study to which the reader is referred for further details [13]. A brief summary of the most relevant concepts is introduced next. The first concept is that of fixed and moving bodies. The fixed body is the contacting body that is conceptually considered to remain fixed in space, while the moving body is the contacting body conceptually considered to move. The position and orientation of the moving body with respect to the fixed body are defined by pose parameters consisting of three translations and three rotations. The second concept is that of sensitive directions. A sensitive direction is a degree of freedom (DOF) which when perturbed causes a relatively large change in associated contact loads. Every sensitive direction possesses an associated sensitive pose parameter and one or more associated sensitive contact loads. For example, if changing the y-translation by a small amount yields a large contact force change in the y-direction, then y-translation is a sensitive pose parameter and y-force is a sensitive load. The third concept is that of a sample point. A sample point is defined as a set of model inputs and corresponding outputs. While one would normally expect to use pose parameters as inputs and loads as outputs, sample point inputs and outputs are permitted to be any combination of pose parameters and loads. For example, sample point inputs could be defined as three rotations and three forces. The fourth concept is use of sample point definitions that contain sensitive pose parameters as outputs and the corresponding sensitive loads as inputs. Sampling in this manner results in a more desirable distribution of sample points since deeply interpenetrating and out-of-contact situations can be avoided. Use of such a sample point definition requires the original model to be sampled via repeated static analyses.

2.2. Multiple domains

In contrast to surrogate modeling methods such as Kriging, feed-forward ANNs can fit tens of thousands of sample points. Therefore, ANNs provide the ability to approximate contact models sampled in a variety of configurations that have not been previously considered. This capability motivates a new sampling approach.

Our sampling strategy consists of combining multiple domains of input space, each with a different span and sample point density. A large span minimizes the likelihood of evaluating the surrogate model outside the sampled domain, a situation that would lead to large prediction errors. A high sample point density leads to low prediction errors within the sampled domain. To maximize model accuracy, we combine sparsely sampled domains having large spans with densely sampled domains having limited spans that cover regions of input space likely to be encountered during the activity being simulated.

To define both types of domains, we introduce the concept of a reference envelope, which we will use to define the upper and lower bound of a domain. We obtain multiple time-histories of a pose parameter or load of interest corresponding to the activity to be simulated. Upper and lower bounds for these curves are defined for each time frame. These bounds comprise the reference envelopes that describe estimated variations in pose parameters and loads of interest across the entire motion.

To define a sparsely sampled domain having a large span, we expand the maximum and minimum values of the reference envelopes across all time frames by user-specified amounts. The resulting time-invariant upper and lower bounds define a domain that forms a large six-dimensional (6-D) hypercuboid input space. This space is filled using a Hammersley quasirandom sequence [21]. The domain should exclude physically unrealistic sections of input space corresponding to contacting surfaces "pulling" on each other.

To define a densely sampled domain having a limited span, we expand the maximum and minimum bounds of the reference envelopes at each time frame by user-specified amounts. The resulting time-varying upper and lower bounds define the domain. For each time frame of the reference motion, we define a 6-D hypercuboid where each dimension corresponds to a sample point input. The 64 vertices of each 6-D hypercuboid are sampled, and the interior of each hypervolume is filled using a 6-D Hammersley sequence (Fig. 1). The resulting sample points will be closely packed and will cluster around the reference envelopes.

Additional domains are added to obtain sample point inputs that place the contact surfaces in an "almost unloaded" condition (contact boundary points) and in an out-of-contact condition. In this way, we include domains to capture specific configurations.