International Journal of Engineering Science 106 (2016) 95–109

Contents lists available at ScienceDirect

International Journal of Engineering Science

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

A technique for the classification of tissues by combining mechanics based models with Bayesian inference

Srikrishna Doraiswamy[a], John C. Criscione[b], Arun R. Srinivasa[a,*]

[a] Department of Mechanical Engineering, Texas A&M University, United States
[b] Department of Biomedical Engineering, Texas A&M University, United States

ARTICLE INFO

Article history:
Received 7 August 2015
Revised 18 January 2016
Accepted 7 April 2016
Available online 2 June 2016

Keywords:
Biomaterials
Bayesian learning
Classification
Variablity
Artery
Constitutive theory

ABSTRACT

This paper deals with a key issue in diagnostics and classification of tissue: given a few samples of tissue which are known to belong to certain categories (such as "healthy" and "diseased") how to categorize a new sample? This is a well known and analyzed problem and very powerful purely data driven approaches have been developed. However, in situations with limited experimental data with a large spread that is typical of biomaterials, a purely data driven approach to classifying the samples is inadequate, since it can suffer from serious bias due to scant data. Here we propose an approach that uses our understanding of the mechanics of the behavior of tissues to transform the problem from the high dimensional space of raw data to a probability distribution on a low dimensional parameter space and then use a Bayesian technique for the classification based on the parameters. A key point in the paper is that the mapping from the raw data to the parameters is not a deterministic mapping (as would be obtained from a lest squares or maximum likelihood approach) but a probabilistic one based on Bayes rule. To apply this rule, there is a need for hypothesis on the prior probability distributions of the parameters themselves and a way to systematically update the hypothesis as more data is obtained. Such a framework should be able to (1) capture prior knowledge that we have about the parameters (such as for example minimum and maximum values for the parameters, likely mean values etc.) ; (2) provide a means for incorporating the knowledge gained from experiments and (3) gradually evolve towards a purely data driven approach as large amounts of data become available.

We utilize a "max-ent" approach to the prior distribution: i.e. we select a distribution that incorporates any available statistical information about the data while being maximally indifferent to all other information. this probability distribution is updated by Bayesian inference where the posterior distributions are obtained by a Markov Chain Monte Carlo (MCMC) sampling method combined with a continuum mechanics based exact solution of a boundary value problem. We illustrate this approach by considering the "soft" classification (i.e we computer the probability of belonging to a class or category) of nominally similar sheep arteries (from two different sheep) based on the probability distribution of

* Corresponding author. Fax: +19792206755.
E-mail addresses: skrishnad@gmail.com (S. Doraiswamy), jccriscione@tamu.edu (J.C. Criscione), asrinivasa@tamu.edu, arun.r.srinivasa@gmail.com (A.R. Srinivasa).

http://dx.doi.org/10.1016/j.ijengsci.2016.04.002

the parameters corresponding to each class This is an alternative to a logistic regression type approach in situations where there is high uncertainty or limited data distribution.

1. Introduction

A central problem of biomechanics is the classification of tissue into different categories based on mechanical response. For example, it is well known that breast cancer tissue has different response from that of normal tissue and this has been used as the basis for elastographic methods of identifying cancer. From the point of view of classification, the problem can be simply stated as follows: given a few sample data (usually referred to as training data) that are known to belong to different categories (such as "'healthy" and "diseased") we need to build a classifier that can assign a category to a new sample. As stated above, the problem does not need any causal model and we can simply use a data driven approach (such as a logistic regression, random forests, k-means clustering etc.) that are now available in textbooks and commercial codes to carry out the classification. These approaches suffer from serious problems when considering tissue mechanics: they work well only if there is sufficient training data available and they cannot be used for extrapolation to other boundary value problems. Furthermore, since the raw data (say for example the stress-strain response) is in very high dimensions, the classifier is not very effective in such a space with the scant data available. However, there is an alternative; we can use our understanding of tissue mechanics to (a) map the raw data into a parameter space using a mechanics model and (b) do the classification in the parameter space. For example, while the exact shape of the stress strain response of breast cancer tissue may change, they are generally stiffer in the hight strain regime. Thus a stiffness parameter can be used for classification. However, there is considerable spread in the data on tissue. One of the biggest challenges for continuum mechanics models that have been proposed to characterize and predict the mechanical response of biomaterials such as blood vessels, the heart and bones (Fung, 1993; Holzapfel & Ogden, 2006; Humphrey, 2002; Mollica, Preziosi, & Rajagopal, 2007) is to identify the model parameters to match experimental data. Conventionally, the least squares approach, wherein the parameters are chosen to minimize the squared error between the model predictions and the experimental data has been used to identify such model parameters. However, unlike metallic/manufactured materials, it is not possible to enforce control in composition, the size and shape of test specimens when considering biomaterials. Neither is it possible to test a large number of samples of closely matching characteristics. Indeed, the variance observed in experiments across different samples reflects this lack of control (see, for example, the data reported in van Andel, Pistecky, and Borst (2003); Carboni, Desch, and Weizsäcker (2007); García-Herrera et al. (2012); Holzapfel, Sommer, Gasser, and Regitnig (2005); Vande Geest, Sacks, and Vorp (2006)). Furthermore, the quality of available experimental data also limits the ability of models in accurately representing and predicting the behavior of biomaterials, i.e. there is substantial epistemic uncertainty in the models.

1.1. Fitting individual sample data with least squares is not useful for diagnostics when we have only a few samples

To show why one needs to look beyond a least squares fit for biomaterials, consider, for example, the data reported by Carboni et al. (2007) where an anisotropic elastic response function with 6-8 constants are used to fit porcine artery data (obtained from pigs from a local abattoir). the values of parameters are chosen to fit each experiment of "nominally" the same body and the results are tabulated in table 2 in their paper. The noticeable feature is the substantial difference in the values of the parameter from one sample to the next in spite of the fact that there are six parameters. Furthermore, even within a single sample, the measured error between the sample response and the model could be as high as 20%.

In spite of this huge uncertainty, such models are useful in characterizing the behavior of such materials especially for classification purposes where the simplicity of a model fitting approach is a significant advantage in spite of the inaccuracies. For example, it has been reported in the literature that elastography studies between normal and diseased liver tissue shows a shear modulus difference is about 3 times. However given the wide spread in the data between samples that are known a-prior to belong to a given group, simple classifiers based on numerical values of parameters obtained through a least squares fit are not reliable. What is needed a probability distribution of parameters for healthy and diseased specimen.

We wish to point out that the problem is two fold; on the one hand, the material properties of tissue vary with a wide range of group characteristics such as age, ethnicity and gender and there is noise associated with this variation. A recent paper (see Seyedsalehi, Zhang, Choi, & Baek (2015)) deals with the group characteristics by invoking a regression based model[1] to account for the variation primarily due to age and develops a technique for using this information as a prior for patient specific modeling. We note that while there are quite a few samples considered by them, they are not all of the same age, i.e. *the number of samples at a given age is very small*.

We seek to address a different problem: Given a few tissue samples at a given age, how to classify them into different categories? *S*amples of tissue from a single individual at a given instant of time also show a huge variation (see Fig. 1).

[1] While classical regression requires specific assumptions about the dependence of the output on the input, Gaussian Process models which allow for non parametric prediction/estimation of the distribution itself and not just point estimates of the mean values, could also be used.