Latent Force Models: Introduction

Neil D. Lawrence

LFM Workshop
13th June 2013
Outline

Motivation

Motion Capture Example
Motivation

Motion Capture Example
Styles of Machine Learning
Background: interpolation is easy, extrapolation is hard

- Urs Hölzle keynote talk at NIPS 2005.
  - Emphasis on massive data sets.
  - Let the data do the work—more data, less extrapolation.
- Alternative paradigm:
  - Very scarce data: computational biology, human motion.
  - How to generalize from scarce data?
  - Need to include more assumptions about the data (e.g. invariances).
General Approach
Broadly Speaking: Two approaches to modeling

- Data modeling
- Mechanistic modeling

Data driven
Knowledge driven
Adaptive models
Differential equations
Digit recognition
Climate, weather models

Figure: Main modeling activity.
General Approach

Broadly Speaking: Two approaches to modeling

- **Data modeling**: let the data “speak”
- **Mechanistic modeling**: impose physical laws
General Approach
Broadly Speaking: Two approaches to modeling

**data modeling**
let the data “speak”

**mechanistic modeling**
impose physical laws
General Approach

Broadly Speaking: Two approaches to modeling

*data modeling*

let the data “speak”
data driven

mechanistic modeling

impose physical laws
General Approach
Broadly Speaking: Two approaches to modeling

- data modeling
  - let the data “speak”
  - data driven

- mechanistic modeling
  - impose physical laws
  - knowledge driven
General Approach
Broadly Speaking: Two approaches to modeling

- data modeling
  - let the data “speak”
  - data driven
  - adaptive models

- mechanistic modeling
  - impose physical laws
  - knowledge driven

Figure: Main modeling activity.
General Approach
Broadly Speaking: Two approaches to modeling

- **data modeling**: let the data "speak"
  - data driven
  - adaptive models

- **mechanistic modeling**: impose physical laws
  - knowledge driven
  - differential equations
General Approach

Broadly Speaking: Two approaches to modeling

<table>
<thead>
<tr>
<th>data modeling</th>
<th>mechanistic modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>let the data “speak”</td>
<td>impose physical laws</td>
</tr>
<tr>
<td>data driven</td>
<td>knowledge driven</td>
</tr>
<tr>
<td>adaptive models</td>
<td>differential equations</td>
</tr>
<tr>
<td>digit recognition</td>
<td></td>
</tr>
</tbody>
</table>

Figure: Main modeling activity.
General Approach
Broadly Speaking: Two approaches to modeling

- data modeling: let the data “speak”, data driven, adaptive models, digit recognition
- mechanistic modeling: impose physical laws, knowledge driven, differential equations, climate, weather models
General Approach
Broadly Speaking: Two approaches to modeling

**data modeling**
- let the data "speak"
- data driven
- adaptive models
- digit recognition

**mechanistic modeling**
- impose physical laws
- knowledge driven
- differential equations
- climate, weather models
General Approach

Broadly Speaking: Two approaches to modeling

*Data modeling*
- let the data "speak"
- data-driven
- adaptive models
- digit recognition

*Mechanistic modeling*
- impose physical laws
- knowledge driven
- differential equations
- climate, weather models

Figure: Main modeling activity.
Underlying data modeling techniques there are weakly mechanistic principles (e.g. smoothness).

In physics the models are typically strongly mechanistic.

In principle we expect a range of models which vary in the strength of their mechanistic assumptions.

Latent Force Models are one part of this spectrum: add further mechanistic ideas to weakly mechanistic models.
Dimensionality Reduction

- Linear relationship between the data, $X \in \mathbb{R}^{n \times p}$, and a reduced dimensional representation, $F \in \mathbb{R}^{n \times q}$, where $q \ll p$.

$$X = FW + \epsilon,$$

$$\epsilon \sim \mathcal{N}(0, \Sigma)$$

- Integrate out $F$, optimize with respect to $W$.

- For Gaussian prior, $F \sim \mathcal{N}(0, I)$
  - and $\Sigma = \sigma^2 I$ we have probabilistic PCA (Tipping and Bishop, 1999; Roweis, 1998).
  - and $\Sigma$ constrained to be diagonal, we have factor analysis.
Dimensionality Reduction: Temporal Data

- Deal with temporal data with a temporal latent prior.
- Independent Gauss-Markov priors over each $f_i(t)$ leads to: Rauch-Tung-Striebel (RTS) smoother (Kalman filter).
- More generally consider a Gaussian process (GP) prior,

$$p(F|t) = \prod_{i=1}^{q} \mathcal{N}(f_{:,i}|0,K_{f_{:,i},f_{:,i}}).$$
Given the covariance functions for \( \{f_i(t)\} \) we have an implied covariance function across all \( \{x_i(t)\} \)—(ML: semi-parametric latent factor model (Teh et al., 2005), Geostatistics: linear model of coregionalization).

- Rauch-Tung-Striebel smoother has been preferred
  - linear computational complexity in \( n \).
  - Advances in sparse approximations have made the general GP framework practical. (Titsias, 2009; Snelson and Ghahramani, 2006; Quiñonero Candela and Rasmussen, 2005).
Mechanical Analogy

Back to Mechanistic Models!

- These models rely on the latent variables to provide the dynamic information.
- We now introduce a further dynamical system with a mechanistic inspiration.
- Physical Interpretation:
  - the latent functions, $f_i(t)$ are $q$ forces.
  - We observe the displacement of $p$ springs to the forces.,
  - Interpret system as the force balance equation, $XD = FS + \epsilon$.
  - Forces act, e.g. through levers — a matrix of sensitivities, $S \in \mathbb{R}^{q \times p}$.
  - Diagonal matrix of spring constants, $D \in \mathbb{R}^{p \times p}$.
  - Original System: $W = SD^{-1}$. 
Extend Model

- Add a damper and give the system mass.

\[ FS = \ddot{X}M + \dot{X}C + XD + \epsilon. \]

- Now have a second order mechanical system.

- It will exhibit inertia and resonance.

- There are many systems that can also be represented by differential equations.
  
  - When being forced by latent function(s), \( \{f_i(t)\}_{i=1}^q \), we call this a latent force model.
Marionette
Mass Spring Damper Analogy

Figure: Mass spring damper analogy, an unobserved force drives multiple oscillators.
Mass Spring Damper Analogy

Figure: Mass spring damper analogy, an unobserved force drives multiple oscillators.
Figure: Mass spring damper analogy, an unobserved force drives multiple oscillators.
Figure: Mass spring damper analogy, an unobserved force drives multiple oscillators.
Mass Spring Damper Analogy

Figure: Mass spring damper analogy, an unobserved force drives multiple oscillators.
Ranked prediction of p53 targets using hidden variable dynamic modeling
Martino Barenco*, Daniela Tomescu*, Daniel Brewer*,†, Robin Callard*,†, Jaroslav Stark‡ and Michael Hubank*†

Addresses: *Institute of Child Health, University College London, Guilford Street, London WC1N 1EH, UK. †CoMPLEX (Centre for Mathematics and Physics in the Life Sciences and Experimental Biology), University College London, Stephenson Way, London, NW1 2HE, UK. ‡Department of Mathematics, Imperial College London, London SW7 2AZ, UK.

Correspondence: Michael Hubank. Email: m.hubank@ich.ucl.ac.uk

Published: 31 March 2006

Received: 24 November 2005
Revised: 30 January 2006
Accepted: 21 February 2006

© 2006 Barenco et al.; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

p53 target prediction
Hidden Variable Dynamic Modelling is a new approach to microarray analysis that quantitatively predicts the regulation of gene activity.

Abstract
Full exploitation of microarray data requires hidden information that cannot be extracted using current analysis methodologies. We present a new approach, hidden variable dynamic modeling (HVDM), which derives the hidden profile of a transcription factor from time series microarray data, and generates a ranked list of predicted targets. We applied HVDM to the p53 network, validating predictions experimentally using small interfering RNA. HVDM can be applied in many systems biology contexts to predict regulation of gene activity quantitatively.

Background
In order to understand how gene networks function, it is necessary to identify their components and to quantitatively describe how they relate to one another [1-3]. Subsequent prediction of gene network behavior requires identification of important parameters and variables, and estimation or measurement of their values during a response [4-6].

Experimental approaches can be applied to identify network components. For example, protein binding arrays and chromosome immunoprecipitation can be applied to identify transcription factor (TF)-binding sites and therefore infer TF targets [7-10]. However, these approaches give a static view of the system. Binding sites identified in vitro may not be available in vivo, and different regulators may be active in different cellular systems. Furthermore, purely experimental approaches cannot predict in a quantitative manner, and with statistical confidence, the dynamics of network activity without making an impractical number of experimental observations [11].

Insight into the dynamic relationships present in a transcriptional response can be gained by running time series of microarrays [3,11,12]. Currently, analysis of this type of datum chiefly relies on clustering or correlation methods. The assumption is that groups of genes with similar expression profiles over time are likely to be regulated by the same TF. Although clustering approaches have been applied with some success, they are limited and inaccurate. Genes with different profiles may still be regulated by the same TF, and many genes included in clusters may be regulated by other factors. Clustering approaches typically do not generate confidence statistics about the validity of individual predictions, and therefore they can neither rank candidates nor distinguish between true and false targets.

Published: 31 March 2006

Received: 24 November 2005
Revised: 30 January 2006
Accepted: 21 February 2006

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2006/7/3/R25
For Gaussian process we can compute the covariance matrices for the output displacements. For one displacement the model is

\[ m_k \ddot{x}_k(t) + c_k \dot{x}_k(t) + d_k x_k(t) = b_k + \sum_{i=0}^{q} s_{ik} f_i(t), \quad (1) \]

where, \( m_k \) is the \( k \)th diagonal element from \( M \) and similarly for \( c_k \) and \( d_k \). \( s_{ik} \) is the \( i,k \)th element of \( S \).

Model the latent forces as \( q \) independent, GPs with exponentiated quadratic covariances

\[ k_{fi,fi}(t, t') = \exp \left( -\frac{(t - t')^2}{2\ell_i^2} \right) \delta_{il}. \]
Covariance for ODE Model

- Exponentiated Quadratic Covariance function for $f(t)$

$$x_j(t) = \frac{1}{m_j \omega_j} \sum_{i=1}^{q} s_{ji} \exp(-\alpha_j t) \int_0^t f_i(\tau) \exp(\alpha_j \tau) \sin(\omega_j(t - \tau)) d\tau$$

- Joint distribution for $x_1(t)$, $x_2(t)$, $x_3(t)$ and $f(t)$.

Damping ratios:

<table>
<thead>
<tr>
<th></th>
<th>$\zeta_1$</th>
<th>$\zeta_2$</th>
<th>$\zeta_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.125</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Covariance for ODE Model

- **Analogy**

\[
x = \sum_{i} e_i^\top f_i \quad f_i \sim \mathcal{N}(0, \Sigma_i) \rightarrow x \sim \mathcal{N}\left(0, \sum_{i} e_i^\top \Sigma_i e_i\right)
\]

- **Joint distribution**
  for \(x_1(t), x_2(t), x_3(t)\) and \(f(t)\).

Damping ratios:

<table>
<thead>
<tr>
<th></th>
<th>(\zeta_1)</th>
<th>(\zeta_2)</th>
<th>(\zeta_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.125</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Covariance for ODE Model

- Exponentiated Quadratic Covariance function for \( f(t) \)

\[
x_j(t) = \frac{1}{m_j \omega_j} \sum_{i=1}^{q} s_{ji} \exp(-\alpha_j t) \int_{0}^{t} f_i(\tau) \exp(\alpha_j \tau) \sin(\omega_j (t - \tau))d\tau
\]

- Joint distribution for \( x_1(t), x_2(t), x_3(t) \) and \( f(t) \).

Damping ratios:

<table>
<thead>
<tr>
<th>( \zeta_1 )</th>
<th>( \zeta_2 )</th>
<th>( \zeta_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.125</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Joint Sampling of $x(t)$ and $f(t)$

Figure: Joint samples from the ODE covariance, black: $f(t)$, red: $x_1(t)$ (underdamped), green: $x_2(t)$ (overdamped), and blue: $x_3(t)$ (critically damped).
Joint Sampling of $x(t)$ and $f(t)$

- lfmSample

**Figure:** Joint samples from the ODE covariance, *black:* $f(t)$, *red:* $x_1(t)$ (underdamped), *green:* $x_2(t)$ (overdamped), and *blue:* $x_3(t)$ (critically damped).
Joint Sampling of $x(t)$ and $f(t)$

- lfmSample

Figure: Joint samples from the ODE covariance, **black**: $f(t)$, **red**: $x_1(t)$ (underdamped), **green**: $x_2(t)$ (overdamped), and **blue**: $x_3(t)$ (critically damped).
Joint Sampling of $x(t)$ and $f(t)$

Figure: Joint samples from the ODE covariance, black: $f(t)$, red: $x_1(t)$ (underdamped), green: $x_2(t)$ (overdamped), and blue: $x_3(t)$ (critically damped).
Exponentiated Quadratic Covariance function for $f(t)$

$$x_j(t) = \frac{1}{m_j \omega_j} \sum_{i=1}^{q} s_{ji} \exp(-\alpha_j t) \int_{0}^{t} f_i(\tau) \exp(\alpha_j \tau) \sin(\omega_j (t-\tau)) d\tau$$

Joint distribution for $x_1(t)$, $x_2(t)$, $x_3(t)$ and $f(t)$.

Damping ratios:

<table>
<thead>
<tr>
<th></th>
<th>$\zeta_1$</th>
<th>$\zeta_2$</th>
<th>$\zeta_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.125</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Outline

Motivation

Motion Capture Example
Example: Motion Capture

Mauricio Alvarez and David Luengo (Álvarez et al., 2009, 2013)

- Motion capture data: used for animating human motion.
- Multivariate time series of angles representing joint positions.
- Objective: generalize from training data to realistic motions.
- Use 2nd Order Latent Force Model with mass/spring/damper (resistor inductor capacitor) at each joint.
Example: Motion Capture

Mauricio Alvarez and David Luengo (Álvarez et al., 2009, 2013)

- Motion capture data: used for animating human motion.
- Multivariate time series of angles representing joint positions.
- Objective: generalize from training data to realistic motions.
- Use 2nd Order Latent Force Model with mass/spring/damper (resistor inductor capacitor) at each joint.
Example: Motion Capture

Mauricio Alvarez and David Luengo (Álvarez et al., 2009, 2013)

- Motion capture data: used for animating human motion.
- Multivariate time series of angles representing joint positions.
- Objective: generalize from training data to realistic motions.
- Use 2nd Order Latent Force Model with mass/spring/damper (resistor inductor capacitor) at each joint.
Example: Motion Capture

- Motion capture data: used for animating human motion.
- Multivariate time series of angles representing joint positions.
- Objective: generalize from training data to realistic motions.
- Use 2nd Order Latent Force Model with mass/spring/damper (resistor inductor capacitor) at each joint.

Mauricio Alvarez and David Luengo (Álvarez et al., 2009, 2013)
Prediction of Test Motion

- Model left arm only.
- 3 balancing motions (18, 19, 20) from subject 49.
- 18 and 19 are similar, 20 contains more dramatic movements.
- Train on 18 and 19 and testing on 20
- Data was down-sampled by 32 (from 120 fps).
- Reconstruct motion of left arm for 20 given other movements.
- Compare with GP that predicts left arm angles given other body angles.
Table: Root mean squared (RMS) angle error for prediction of the left arm’s configuration in the motion capture data. Prediction with the latent force model outperforms the prediction with regression for all apart from the radius’s angle.

<table>
<thead>
<tr>
<th>Angle</th>
<th>Latent Force Error</th>
<th>Regression Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radius</td>
<td>4.11</td>
<td>4.02</td>
</tr>
<tr>
<td>Wrist</td>
<td>6.55</td>
<td>6.65</td>
</tr>
<tr>
<td>Hand X rotation</td>
<td>1.82</td>
<td>3.21</td>
</tr>
<tr>
<td>Hand Z rotation</td>
<td>2.76</td>
<td>6.14</td>
</tr>
<tr>
<td>Thumb X rotation</td>
<td>1.77</td>
<td>3.10</td>
</tr>
<tr>
<td>Thumb Z rotation</td>
<td>2.73</td>
<td>6.09</td>
</tr>
</tbody>
</table>
Mocap Results II

(a) Inferred Latent Force

(b) Wrist

(c) Hand X Rotation

(d) Hand Z Rotation

(e) Thumb X Rotation

(f) Thumb Z Rotation

Figure: Predictions from LFM (solid line, grey error bars) and direct regression (crosses with stick error bars).
Data set is from the CMU motion capture data base\textsuperscript{1}.

Two different types of movements: golf-swing and walking.

Train on a subset of motions for each movement and test on a different subset.

This assesses the model’s ability to extrapolate.

For testing: condition on three angles associated to the root nodes and first five and last five frames of the motion.

Golf-swing use leave one out cross validation on four motions.

For the walking train on 4 motions and validate on 8 motions.
## Motion Capture Results

**Table:** RMSE and $R^2$ (explained variance) for golf swing and walking

<table>
<thead>
<tr>
<th>Movement</th>
<th>Method</th>
<th>RMSE</th>
<th>$R^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Golf swing</td>
<td>IND GP</td>
<td>$21.55 \pm 2.35$</td>
<td>$30.99 \pm 9.67$</td>
</tr>
<tr>
<td></td>
<td>MTGP</td>
<td>$21.19 \pm 2.18$</td>
<td>$45.59 \pm 7.86$</td>
</tr>
<tr>
<td></td>
<td>SLFM</td>
<td>$21.52 \pm 1.93$</td>
<td>$49.32 \pm 3.03$</td>
</tr>
<tr>
<td></td>
<td>LFM</td>
<td><strong>18.09 \pm 1.30</strong></td>
<td><strong>72.25 \pm 3.08</strong></td>
</tr>
<tr>
<td>Walking</td>
<td>IND GP</td>
<td>$8.03 \pm 2.55$</td>
<td>$30.55 \pm 10.64$</td>
</tr>
<tr>
<td></td>
<td>MTGP</td>
<td>$7.75 \pm 2.05$</td>
<td>$37.77 \pm 4.53$</td>
</tr>
<tr>
<td></td>
<td>SLFM</td>
<td>$7.81 \pm 2.00$</td>
<td>$36.84 \pm 4.26$</td>
</tr>
<tr>
<td></td>
<td>LFM</td>
<td><strong>7.23 \pm 2.18</strong></td>
<td><strong>48.15 \pm 5.66</strong></td>
</tr>
</tbody>
</table>


