MoDisc User's Manual

Ranjan Srivastava

Last modified - March 1, 2006

Contents

Chapter 1

Introduction

Model development is a useful tool for understanding a wide range of phenomena in the sciences and engingeering. Within the area of biological sciences, this approach is becoming more important as scientists strive to keep up with the rapid influx of data being generated. In attempting to make sense of biological phenomena, several models of a system may be postulated. The question then becomes how to discriminate among the models to determine which is most likely.

A method for identifying the most probable model of a chemical reaction network based on experimental data was developed by Stewart and colleagues [2, 3]. This approach, termed “model discrimination,” is a Bayesian-based method in which the probability of a model, given a set of experimental data, may be calculated and compared to other potential models. The model with the highest value is considered the most probable model to describe the system.

Chapter 2

Model Discrimination Theory

Stewart's Method of model discrimination is based upon Bayesian analysis. A full derivation of the technique is provided in [2] and [3]. However a brief description of the principal points are provided here.

Stewart's Method is based on calculating the posterior probabilities of competing models relative to each other. The model with highest probability is considered the most likely choice relative to the other models. Calculation of posterior probabilities are based on the following proportionality,

$$p(M_j \mid \boldsymbol{Y}) \propto p(M_j) 2^{-\frac{p_j}{2}} \mid \hat{\mathbf{v}}_{\boldsymbol{j}} \mid^{-\frac{\nu_e}{2}} \tag{2.1}$$

where M_j is the j^{th} model, $\boldsymbol{Y}$ is the matrix of experimental results, and $p(M_j \mid \boldsymbol{Y})$ is the posterior probability of model M_j given $\boldsymbol{Y}$. Additionally, p_j is the number of independent parameters, ν_e is the number of degrees of freedom, and $\mid \hat{\mathbf{v}}_{\boldsymbol{j}} \mid$ is the determinant function. The elements of the determinant function are given by

$$v_{ik}(\boldsymbol{\theta_j}) = \sum_{u=1}^{n} [Y_{iu} - \mathcal{F}_{ji}(\xi_u, \boldsymbol{\theta_j})][Y_{ku} - \mathcal{F}_{jk}(\xi_u, \boldsymbol{\theta_j})] \qquad i, j = 1, \ldots, q \tag{2.2}$$

in which n is the the total number of events evaluated, q is the number of different chemical species monitored, and $\mathcal{F}_{ji}(\xi_u,\boldsymbol{\theta_j})$ is the model prediction. $\boldsymbol{\xi}$ is the vector of the number of different independent conditions (i.e. temperature) tested, and $\boldsymbol{\theta_j}$ is the vector of parameters providing the best fit of the model to the experimental data.

Normalizing the results of Equation 2.1 for any given model to the sum of the results for all the models is referred to as the *probability share* and is shown in Equation 2.3,

$$\pi_j(M_j \mid \boldsymbol{Y}) = \frac{p(M_j \mid \boldsymbol{Y})}{\sum_k p(M_k \mid \boldsymbol{Y})} \tag{2.3}$$

The model with the highest probability share is considered most probable.

Chapter 3

MoDisc Usage

3.1 Software Installation

To carry out installation of MoDisc, it is first necessary to download and install the LispWorks Common Lisp Personal Edition software. The software may be freely downloaded from http://www.lispworks.com/downloads/index.html, along with documentation of how to install the software. LispWorks is available for Windows, Linux, and OS X.

Once the LispWorks Personal Edition is installed, the MoDisc code may be downloaded. MoDisc is available at http://common-lisp.net/project/modisc/. The link for the software is found on the left-hand side menu bar. You may also directly download it from http://common-lisp.net/project/modisc/modisc.zip.

3.2 Input File

To use MoDisc, experimental data and model information may be entered via an input file. The file may be in the form of a spreadsheet, such as an Excel file or as a tab delimited text file. The input file is organized into a series of blocks of information for MoDisc.

The first block is for the comments sections. Comments may be entered within a “begin-comment” and “end-comment” section.

The next set of of blocks are for the details of the model, as well as the results of the simulation. Model simulation results may be entered by starting a section called “begin-model” followed by the name of the model. The number of parameters, the number of variables, and the number of degree of freedoms are entered next. The following row consists of a list of the variables used in the simulation. Simulation results for the specific model are then entered. Finally the block is closed by ending it with an “end-model” row. This procedure is repeated for each of the remaining models.

To enter the experimental data, a new block is started by entering “begin-experiment” in a new row. The number of variables are then entered, followed by a row consisting of the independent and dependent variables measured. Finally the experimental data is entered, followed by an “end-experiment” row. The file may then be saved as a tab delimited text file.

3.3 Running MoDisc

To run MoDisc, first launch LispWorks. Under the LispWorks menu bar, choose “File $\rightarrow$ Open” and select the “modisc.lsp” file. This will result in an editor window being launched containing the modisc source code. Select the editor window using your mouse. Then go to the menu bar and select “Buffers $\rightarrow$ Compile.” Finally, at the “CL-USER 1 $>$” prompt in the original LispWorks window, type “(modisc)” (make sure to include the parenthesis). At this point, you will be prompted for your input file. After reading the input file in, MoDisc will return the probability share of each of the models.

Chapter 4

Examples

4.1 HIV-1 Viral Dynamics

Details regarding the results of model discrimination for several deterministic models of HIV-1 viral dynamics have been reported in the journal *Bioinformatics* [1]. Details may be found at http://bioinformatics.oxfordjournals.org/cgi/content/abstract/21/8/1668.

4.2 MS2 Bacteriophage Viral Dynamics

Coming soon.

4.3 Insulin-Glucose Dynamics

Coming soon.

4.4 Identification of Most Probable Objective Function for Flux Balance Analysis

Coming soon.

Chapter 5

Still To Do

2/26/06 - Over the course of the next few weeks, the following items will be taken care of:

- Generate binaries for different platforms to make MoDisc easier to launch.
- Add figures to “Running MoDisc” subsection to help explain launching procedure.
- Clean up source code.
- Add comments to source code.
- Flesh out model discrimination examples in “Examples” chapter.
- Provide example file.

Bibliography

[1] A. L. Knorr and R. Srivastava. Evaluation of HIV-1 kinetic models using quantitative discrimination analysis. *Bioinformatics*, 21(8):1668–77, 2005.

[2] W.E. Stewart, T.L. Henson, and G.E.P. Box. Model discrimination and criticism with single-response data. *AIChE Journal*, 42(11):3055–3062, 1996.

[3] W.E. Stewart, Y. Shon, and G.E.P. Box. Discrimination and goodness of fit of multiresponse mechanistic models. *AIChE Journal*, 44(6):1404–1412, 1998.