Selection of Perturbation Experiments for Model Discrimination

Ivayla Vatcheva, Hidde de Jong and Nicolaas J.I. Mars

Abstract. It often occurs that a system can be described by several competing models. In order to distinguish among the alternative models, further information about the behavior of the system is required. One way to obtain such information is to perform suitably chosen perturbation experiments. We introduce a method for the selection of optimal perturbation experiments for discrimination among a set of dynamical models. The models are assumed to have the form of semi-quantitative differential equations. The method employs an optimization criterion based on the entropy measure of information.

1 Introduction

Scientists and engineers are frequently faced with situations in which a given system can be described by several competing models. When analyzing the synthesis rate of a product in a catalyzed chemical reaction, one may end up with several equations that all satisfy a set of measurements [13]. For the mitotic clock of early embryos, a dozen of models predicting the observed periodic behavior of the concentrations of key proteins has been suggested [9].

In order to identify which of the given models best describes the actual situation, new observations have to be made. These can be obtained by performing supplementary perturbation experiments. In a perturbation experiment the structure of the system and/or the experimental conditions are changed. An experiment discriminates between the competing models, if the predictions of some of the candidates, which have been properly modified to reflect the experimental change, fit the newly obtained data whereas others do not. The problem of model discrimination can then be defined as the problem of selecting an experiment that gives rise to observations matching the predictions of as few of the models as possible.

Incomplete knowledge about the system to be understood does not always permit the formulation of detailed quantitative models. In what follows we assume that the models are given in the form of semi-quantitative differential equations. Appropriate semi-quantitative simulation techniques are used to derive interval predictions for the model variables. As measurements obtained in the experiments may be imprecise, they are considered to be intervals as well.

We present a method for the systematic choice of perturbation experiments for the discrimination of semi-quantitative dynamical models. Experiments are selected on the basis of a generalization of an entropy criterion suggested by Box & Hill [2], which measures the information increment provided by each experiment. The concept of entropy as a discrimination criterion has also been used in statistics (e.g. [10]), and in model-based diagnosis (e.g. [4, 12]). A novel aspect of our work is that we extend this concept to the case of perturbation experiments and to situations in which experimental systems are described by semi-quantitative dynamical models.

The in-principle applicability of our approach is illustrated on a set of competing models of an oscillatory, second-order system. We will consider six models of a mass-spring system and illustrate the choice of suitable perturbations to discriminate between the models. The principles involved in this example are applicable to the investigation of more complex and less understood oscillating systems.

The presentation starts with a description of the problem of model discrimination. A number of basic concepts are introduced and the relationship between models and experiments is given. The criterion of model discrimination can then be defined as the problem of selecting an experiment that gives rise to observations matching the predictions of as few of the models as possible.

Suppose a set $M$ of models of an experimental system has been proposed. Let $p(m_i)$ be the a priori probability of $m_i \in M$ being the correct model of the system. The probabilities can be derived from preliminary observations on the system behavior or theoretical considerations. If no prior knowledge about the relative plausibilities of the models exists, equal probabilities are assumed. We say that the models in $M$ are competing. $M$ is assumed to be complete, that is, $\sum_{m_i \in M} p(m_i) = 1$.

In this paper we will model experimental systems by means of semi-quantitative differential equations (SQDEs), that is, qualitative differential equations (QDEs) enhanced with numerical information. The quantitative information completing a QDE takes the form of numerical ranges added to landmarks and of envelopes for monotonic function constraints [1]. Fig. 1 shows two SQDEs describing a simple mass-spring system. The models assume that the forces playing a role in the experiment are a spring force and a friction force, but they differ in the precise nature ascribed to the former.

1 Department of Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands, email: {ivayla.mars}@cs.utwente.nl
2 Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de Recherche Rhône-Alpes, 655 avenue de l’Europe, 38330 Montbonnot Saint-Martin, France, email: Hidde.de-Jong@inrialpes.fr
Predicted behavioral features need to correspond with observed behavioral features of the system. That is, it should be possible to relate a predicted behavioral feature to some direct or indirect measurement of quantities of the system. As measurements will be assumed to have the form of confidence intervals, observed behavioral features are intervals.

The results of a perturbation experiment can be used to recompute the probabilities of the competing models. Models of which the predictions do not agree with the observations will have an a posteriori probability equal to 0. The model discrimination problem can now be intuitively stated as follows: find the experiment with values for the observed behavioral features that make a maximum number of models improbable. In the next section, we elaborate this intuition by means of an approach based on concepts from information theory.

3 Method for the selection of perturbation experiments

We will be interested in finding the perturbation yielding the highest increment in information [2]. Consider a behavioral feature $Y$, with
interval values in $D \subseteq \mathbb{R}$. Let $e \in E$ be a perturbation experiment, whose outcome yields a value $Y^e = [y^e - \epsilon/2, y^e + \epsilon/2]$ of the behavioral feature, where $y^e$ is the midpoint of the interval $Y^e$ and $\epsilon$ is the size of the confidence interval for $Y$. The information increment of $e$ is formulated as

$$\Delta H(e) = - \sum_{m_i \in M} p(m_i) \ln p(m_i) + \sum_{m_i \in M} p(m_i \mid Y^e) \ln p(m_i \mid Y^e),$$

(1)

where $p(m_i)$ and $p(m_i \mid Y^e)$ are the a priori and a posteriori probabilities of $m_i$. $\Delta H$ reaches its maximum when the a posteriori probabilities of all models but one are 0. A minimal value is attained when the a posteriori probabilities are equal.

The $p(m_i \mid Y^e)$ is in (1) are not known, since they are determined by the outcome of the experiment. However, we can express the expected value of $\Delta H$ in terms of the probability distributions $g_i^e(Y^e)$ of the behavioral feature $Y$. For brevity, $g_i^e$ instead of $g_i^e(Y^e)$ will be used if no confusion about the behavioral feature being considered is possible. The value of $Y$ predicted by $m_i$ under perturbation $e$ is an interval $V_i^e \subseteq D$, with distribution $g_i^e : D \rightarrow \mathbb{R}_{\geq 0}$ defined as follows

$$g_i^e(y) = \begin{cases} \frac{[y - \frac{\epsilon}{2}, y + \frac{\epsilon}{2}] \cap V_i^e}{|V_i^e|}, & y - \frac{\epsilon}{2}, y + \frac{\epsilon}{2} \cap V_i^e \neq \emptyset, \\ 0, & y - \frac{\epsilon}{2}, y + \frac{\epsilon}{2} \cap V_i^e = \emptyset, \end{cases}$$

(2)

where $|\cdot|$ denotes an interval length. $g_i^e(y)$ expresses the probability that the empirically-determined value of $Y$ is $[y - \epsilon/2, y + \epsilon/2]$ if model $m_i$ is the correct model. (2) can be replaced by the following equivalent expression, where the $g_i^e$s are defined as piecewise-linear functions:

$$g_i^e(y) = \begin{cases} \frac{y - V_i^e + \epsilon/2}{|V_i^e|}, & y \in [V_i^e - \epsilon/2, V_i^e + \epsilon/2], \\ \frac{y - V_i^e - \epsilon/2}{|V_i^e|}, & y \in [V_i^e + \epsilon/2, V_i^e - \epsilon/2], \end{cases}$$

(3)

and $V_i^e$ and $V_i^e$ denote the lower and the upper bound of $V_i^e$, respectively. Fig. 4 illustrates the function $g_i^e$ for an experiment $e$ consisting of replacing the object in the mass-spring system by a lighter object. The behavioral feature in this case is the interval value for the amplitude of the system.

![Figure 4](image)

A plot of the function $g_i^e$ for the amplitude of the mass-spring system in an experiment consisting of replacing the object by a lighter object ($e_4$ in the next section). The prediction of model $m_3$ (Fig. 6) perturbed according to this experiment is $V_{3e} = [0.46, 0.67]$, and $e = 0.1$.

Call the expected value of the information increment $\Delta J(e)$. By definition,

$$\Delta J(e) = \int_{y \in D} \Delta H(e) g_i^e(y) dy,$$

(4)

where $g_i^e(y) = \sum_{m_i \in M} p(m_i) g_i^e(y)$. By substituting the expression for $\Delta H(e)$ in (4) we get,

$$\Delta J(e) = \sum_{m_i \in M} p(m_i) \int_{y \in D} g_i^e(y) \left( \sum_{m_j \in M} p(m_j \mid Y) \ln p(m_j \mid Y) - \sum_{m_j \in M} p(m_j) \ln p(m_j) \right) dy,$$

(5)

where $Y = [y - \epsilon/2, y + \epsilon/2]$ and

$$p(m_j \mid Y) = \frac{p(m_i) g_j^e(y)}{g_i^e(y)}$$

(6)

via the Bayes rule. Combination of (5) and (6) gives, after algebraic simplification,

$$\Delta J(e) = \sum_{m_i \in M} p(m_i) \int_{y \in D} g_i^e(y) \ln \frac{g_i^e(y)}{g_j^e(y)} dy.$$
\[ QV(a) = -\frac{QV(g)}{QV(l)} QV(x) - \frac{QV(c)}{QV(m)} QV(|v|) QV(v) \]

\[ (m_3) \]

\[ QV(a) = -\frac{QV(g)}{QV(l)} \left( QV(x) + \frac{QV(x^2)}{QV(k)} \right) - \frac{QV(c)}{QV(m)} QV(v) \]

\[ (m_4) \]

\[ QV(a) = -\frac{QV(g)}{QV(l)} \left( QV(x) + \frac{QV(x^2)}{QV(k)} \right) - \frac{QV(c)}{QV(m)} QV(|v|) QV(v) \]

\[ (m_5) \]

\[ QV(a) = -\frac{QV(g)}{QV(l)} \left( QV(x) + \frac{QV(x^2)}{QV(k)} \right) - \frac{QV(c)}{QV(m)} QV(|v|) QV(v) \]

\[ (m_6) \]

**Figure 6.** Models \( m_3-m_6 \) together with \( m_1 \) and \( m_2 \) in Fig. 1 form a set of competing models for the mass-spring system. Models \( m_1 \) and \( m_3 \) assume linear spring force. Models \( m_2 \) and \( m_4 \) assume soft spring forces (the stiffness of the spring decreases with the displacement), while the spring force in \( m_5 \) and \( m_6 \) is hard (the stiffness increases with the displacement). In \( m_5 \), \( m_2 \) and \( m_3 \) the acceleration depends linearly on the velocity. The models \( m_5 \), \( m_4 \) and \( m_6 \) assume quadratic dependency. The meaning of the variables and the constraints for \( QV(x) \) and \( QV(v) \) is the same as in Fig. 1.

\[ \begin{array}{c|ccccc|c|ccccc} m_1 & e_1 & e_2 & e_3 & e_4 & e_5 & \Delta J(e_1) \\ \hline 0.9, 1.1 & 0.12, 0.39 & 0.8, 1.01 & 0.63, 0.97 & 1.13, 1.48 & 0 & e_1 \\ m_2 & 0.9, 1.1 & 0.11, 0.38 & 0.81, 1.01 & 0.62, 0.88 & 1.57, 2.02 & 0.5315 \\ m_3 & 0.9, 1.1 & 0.29, 0.52 & 0.88, 1.01 & 0.46, 0.67 & 1.24, 1.65 & 0.0934 (b) \\ m_4 & 0.9, 1.1 & 0.34, 0.55 & 0.8, 1.0 & 0.41, 0.62 & 1.22, 1.64 & 0.7499 \\ m_5 & 0.9, 1.1 & 0.18, 0.43 & 0.81, 1.01 & 0.64, 0.86 & 1.34, 1.77 & 0.6504 \\ m_6 & 0.9, 1.1 & 0.23, 0.51 & 0.83, 1.02 & 0.41, 0.62 & 1.09, 1.47 & 0.1694 \\ \end{array} \]

**Table 1.** (a) Predictions for feature \( f_1 \), the interval value for the amplitude derived from the models for the perturbations \( q_1, \ldots, q_6 \). (b) The values of \( \Delta J \) computed for all perturbations, and (c) some values of \( \Delta J \) after application of \( q_6 \) (see text).

\( p(m_j) \) be the a priori probabilities of the models and \( E \) a set of predefined perturbation experiments.

while (\( \exists m_i \in M : p(m_i) \neq 0 \) and \( \forall m_i \in M : p(m_i) < \theta \) and not \( E \) empty) do

\( \text{determine } e \in E \text{ for which } \Delta J(e) \text{ is maximal} \)

\( \text{perform experiment corresponding to } e, \text{ determine } Y^e \)

\( \text{compute the a posteriori probabilities } p(m_j | Y^e) \)

\( \text{set } p(m_j) \text{ to } p(m_j | Y^e) \)

\( \text{remove } e \text{ from } E \)

The algorithm selects perturbation experiments until one of the following happens: a model has a sufficiently high probability, all models have zero probabilities, or all possible experiments have been executed. If the algorithm terminates with \( p(m_i) = 0 \) for all models, obviously the assumption for completeness of \( M \) is violated.

### 4 Example and evaluation

Consider the six models of a mass-spring system listed in Fig. 1 and Fig. 6 [11]. The models differ in the terms for the spring and the friction force. The experiment consists in stretching and then releasing the spring. Assume the following perturbation experiments can be performed. \( e_1 \): replace the medium by an approximately frictionless medium (\( c = 0 \)); \( e_2 \): replace the medium by a more compact medium (\( c = [2.85, 3.15] \)); \( e_3 \): test with a heavier object having mass \([11.95, 12.05]\); \( e_4 \): test with a lighter object having mass \([0.7, 0.8] \); \( e_5 \): release the object with initial velocity \([1.9, 2.2] \).

We consider four behavioral features: \( Y_1 \) is the interval value of the maximum distance from the rest position (the amplitude); \( Y_2 \) is the interval value of the frequency of the system; \( Y_3 \) is the relative interval value of the maximum amplitude for the perturbed and the original system; and \( Y_4 \) is the relative interval value of the frequency.

Values for \( Y_1 \) to \( Y_4 \) have been derived from the perturbed models by means of semi-quantitative simulation and comparative analysis. The predicted intervals for the amplitude are shown in Table 1(a). The first perturbation gives rise to identical predictions from all models. It is evident, even without looking at the value of \( \Delta J(e_1) \), that the corresponding experiment will never distinguish between the models. The rest of the perturbations also do not give distinct intervals for this feature, but the predictions are not entirely overlapping. Hence, the measurements in the corresponding experiments may discriminate between at least some of the models.

Assume the amplitude of the system to be the only quantity measured in the experiments. Suppose the models have equal a priori probabilities \( p(m_1) = \ldots = p(m_6) = 1/6 \) and \( \theta = 0.75 \), that is, a model is considered best if its probability is larger than 0.75. At the first step of the algorithm, \( e_4 \) is chosen since it maximizes \( \Delta J \) (see Table 1(b)). Assume the experiment is executed and a measurement \([0.4207, 0.5207]\) is obtained. The measurement is not consistent with the predictions derived from \( m_1, m_2, \) and \( m_5 \) for this perturbation, so that the a posteriori probabilities of these models become 0. The a posteriori probabilities of the other three models after the experiment are \( p(m_3) = 0.2330 \), \( p(m_4) = 0.3835 \) and \( p(m_6) = 0.3835 \). In the next iteration, \( e_2 \) is selected (Table 1(c)). Assume the measurement \([0.3080, 0.4080]\) is obtained which gives rise to the posterior probabilities \( p(m_3) = 0.2794 \), \( p(m_4) = 0.3428 \) and \( p(m_6) = 0.3778 \). Next, \( e_5 \) is chosen. A measurement \([1.1340, 1.2340]\) causes \( p(m_3) = 0 \) and the algorithm terminates, giving \( m_5 \) as the best model of the system with \( p(m_5) = 0.8964 \).

In order to evaluate the performance of the method we have adopted the following strategy. First, one of the models (\( m_6 \)) was arbitrary selected. “Experimental” data was then produced by generating random intervals within the predictions of \( m_6 \). The length of the random intervals was set equal to the size of the confidence interval of the behavioral feature (\( c = 0.1 \) in the case of the amplitude). Finally, the algorithm of the previous section was applied given these data. This procedure was repeated 20 times and the results analyzed.

In only 15% of the cases the model that was used to generate the data was identified as the single remaining candidate. In the rest of the cases the algorithm terminated with two to three candidate models that could not be discriminated. On average, for the identification of the model 4 experiments were necessary. For comparison, when the size of the confidence interval was taken to be 0.01, in 40% of the cases \( m_6 \) was identified with average number of experiments 2.5. The results show, not surprisingly, that when the measurement error is smaller, better discrimination is achieved.

Now suppose all four features are considered, the other circum-
stices remaining the same. In this case, $e_5$ maximizes $\Delta J$ and it is selected as the best experiment (see the table below). Values of $[1,134, 1,234]$ and $[4,12, 4,22]$ for the amplitude and the period of the perturbed system, for instance, give rise to the posteriori probabilities $p(m_1) = \ldots = p(m_5) = 0$ and $p(m_6) = 1.0$.

<table>
<thead>
<tr>
<th>$\Delta J(e_i)$</th>
<th>$e_1$</th>
<th>$e_2$</th>
<th>$e_3$</th>
<th>$e_4$</th>
<th>$e_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.4049</td>
<td>1.1429</td>
<td>1.5548</td>
<td>1.5285</td>
<td>3.1811</td>
</tr>
</tbody>
</table>

The above evaluation procedure was again applied 20 times, now for the situation that all four features are taken into account. We found that the average number of experiments necessary to identify model $m_0$ was 1.1. In only two of the cases a second iteration in the algorithm was necessary. In all cases complete discrimination was achieved.

The example illustrates that when more behavioral features are considered, a higher efficiency may be achieved: measuring only the amplitude, we needed four experiments to discriminate between the models, while taking into account all four behavioral features a single experiment turned out to be sufficient.

Evaluation by means of random data was used to investigate the performance improvement of the algorithm for experiment selection with respect to random selection of perturbation experiments. Assume the amplitude of the system is the only quantity being measured ($\epsilon = 0.1$). After 20 times we again obtained that 4 experiments are necessary, on average, to identify the correct model. The reasons for the lack of improvement of our method with respect to random selection are the large overlap between (some of) the predictions, the high measurement error assumed, and the low number of experiments provided. However, selecting the experiments in random order when all four features were considered, required 3.2 experiments on average to identify the correct model, whereas selecting the experiments by our method required only 1.1 experiments.

5 Discussion and conclusion

We have presented a method for discriminating among competing semi-quantitative models by selecting suitable perturbation experiments. The method chooses a maximally-discriminating experiment by means of a criterion based on the entropy measure of information. The application of this criterion was illustrated in an example concerning a set of competing models of a mass-spring system.

Information theory has been used in model-based diagnosis to distinguish among competing diagnoses of a faulty system (e.g. [4]). Like our method, these methods proceed by making new observations on the system. However, the work mentioned above is limited to determining the best measurement point within a given experiment, while we seek the best experiment that would permit optimal discrimination. Struss [12] has extended the approach in [4] by finding the best operating conditions that would give rise to the most discriminatory observations. Our work attempts to generalize this method by employing dynamical models and by extending the concept of discriminating test to discriminating perturbation experiment.

In statistics, the idea of employing the entropy measure as a discrimination criterion has been illustrated for distinguishing between quantitative algebraic models (e.g. [2, 10]). In [7] the entropy has been used to design observations discriminating among rival water quality models. However, these examples are restricted to fully numerical models with precise point measurements. In this paper, we have shown how the criterion can be generalized to the case that only imprecise, approximate observations of the system are available.

The idea of planning perturbation experiments for model discrimination based on an entropy measure has also been proposed in [6]. Models of a genetic regulation network are discriminated by varying the expression level of involved genes or the influence of external stimuli. This method, however, is limited to models in the form of Boolean networks and to binary perturbations. This article generalizes the approach in [6] by employing more advanced dynamical models and by extending the concept of perturbation experiments.

The work presented here can be extended into several directions. In practice, the number of possible perturbations will be infinite when the value of a quantity can be changed continuously. The problem of model discrimination as defined here should then be generalized. Instead of selecting a discrete perturbation that has been specified beforehand, a value for the quantity that maximizes ($7$) has to be chosen. An issue neglected thus far are the costs associated with experiments. In practice, the costs for performing an experiment may need to be balanced against its expected utility. In these cases, the problem can be reformulated as the selection of an experiment that maximizes $\Delta J(e) / h(\epsilon(e))$, where $h$ is a function depending on the intended application: one may be interested in effective experiments without caring about expenses, or prefer less costly tests.

Further research will concentrate on the extension of the method along the lines mentioned above, its comparison with other model discrimination techniques (e.g. [5]), and its application to real-world systems. Currently, we are applying the approach to a model discrimination problem in biology: the regulation of the cell cycle in early embryos [9]. This system is described by second-order models and exhibits periodic behavior similar to the oscillations of the mass-spring example considered here.

REFERENCES