
Equation Discovery for Model Identification in Respiratory Mechanics under Conditions of Mechanical Ventilation

Steven Ganzert

STEVEN.GANZERT@UNIKLINIK-FREIBURG.DE

Dept. of Anaesthesiology and Critical Care Medicine, University Hospital Freiburg, Hugstetter Str. 55, 79106 Freiburg, Germany

Knut Möller

MOE@HS-FURTWANGEN.DE

Dept. of Biomedical Engineering, University of Applied Sciences, Jakob Kienzle Str. 17, 78054 Villingen-Schwenningen, Germany

Kristian Kersting

KERSTING@CSAIL.MIT.EDU

CSAIL, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, MA 02139, USA

Luc De Raedt

LUC.DERAEDT@CS.KULEUVEN.BE

Dept. of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium

Josef Guttmann

JOSEF.GUTTMANN@UNIKLINIK-FREIBURG.DE

Dept. of Anaesthesiology and Critical Care Medicine, University Hospital Freiburg, Hugstetter Str. 55, 79106 Freiburg, Germany

Abstract

Lung protective ventilation considerably improves the outcome of mechanically ventilated and critically ill patients as it avoids extensive mechanical stress of the lung tissue and hence its irreversible damage. A valid analysis of respiratory mechanics is a prerequisite for lung protective ventilation. This analysis is always based on mathematical models. The equation of motion defines a generally accepted model of the respiratory system. It relates the airway pressure to the ventilator induced airflow and volume application influenced by distensibility and resistance of the respiratory system. We present a novel equation discovery system which combines the technique of using declarative bias for the reduction of the vast search space known from the LAGRAMGE-system with a greedy, randomized search strategy according to GSAT. We experimentally validate the effectiveness of our approach and show that the equation of motion model can automatically be rediscovered from real-world data.

Appearing in working notes of the ICML07 workshop on the *Induction of Process Models*, June 24, Corvallis, OR, USA, 2007. Copyright 2007 by the author(s)/owner(s).

1. Introduction

Mechanical ventilation is on the one hand the livesaving therapy in intensive care medicine by all means. On the other hand this therapy can even aggravate the pulmonary status of the critically ill patient. Mechanical damage of the lung tissue is the predominant reason for ventilator induced lung injury (VILI). To avoid VILI, the physicians make great efforts to develop lung protective ventilation strategies. The equation of motion (EOM) is the commonly accepted mathematical model of the respiratory system which provides the basis for the most clinically applied methods of respiratory mechanics analysis.

The aim of the present study was to show that the EOM can be automatically discovered from real-world data using a novel equation discovery system. The system combines the use of a strong declarative bias to reduce the search space complexity with a novel search heuristic for traversing the resulting hypothesis space. This search heuristic uses a greedy, randomized strategy according to the GSAT algorithm. The technique of using declarative bias in order to reduce the hypothesis space is well known from the field of ILP (inductive logic programming) (DeRaedt, 1996) and has been introduced to the field of equation discovery by the LAGRAMGE-system (Todorowski & Džeroski, 1997).

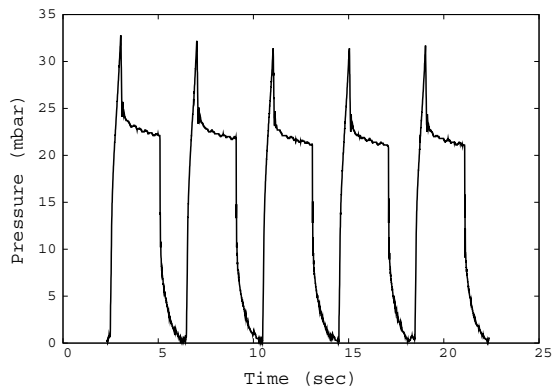


Figure 1. Example for a typical real-world pressure time series.

In several experiments, including standard tests from the field of equation discovery and simulated data in a medical problem domain, we empirically show the effectiveness of our approach. Furthermore, we empirically evaluate our method on respiratory data of mechanically ventilated patients, suffering from the ARDS (adult respiratory distress syndrome) in order to test the capability of the system to reidentify the EOM. The latter would be the methodological prerequisite to be able to identify even more elaborated models for lung mechanics analysis.

We proceed as follows: After briefly reviewing the medical background with respect to the EOM, the novelties in our equation discovery approach are introduced. Before concluding, we present and discuss the experiments carried out to assess the performance of the system.

2. Medical Background

The EOM describes the relation between the airway pressure P_{aw} (cf. Figure 1), the volume Vol and the airflow $Flow$ during artificial ventilation of the human lung:

$$P_{aw} = Vol/C + Flow * R + PEEP.$$

The variable $PEEP$ (Positive End Expiratory Pressure) indicates the airway pressure at the end of expiration. C indicates the volume distensibility and R the resistance of the respiratory system.

3. Equation Discovery

The presented equation discovery approach essentially follows the LAGRAMGE-system (Todorowski & Džeroski, 1997). The input to the system consists of

two parts. Firstly the data set, consisting of one or more time series datafiles, each again including one or more parameter variable records, cf. Fig 1.

Secondly a grammar, specifying the declarative bias, cf. Fig 2: The productions are ordered by the minimal depth of a deriveable parse tree. Productions with the same left side are gathered in one single production rule, separated by '|'. Each deriveable parsetree represents a possible model. Preferring short hypothesis, the system generates parse trees in the sequence of this order, i.e in a left-to-right manner.

$$\begin{array}{l|l|l} E & \rightarrow & E + F & | & E - F & | & F \\ F & \rightarrow & E * T & | & F/T & | & T \\ T & \rightarrow & const[:0 : 25 : 70] & | & V & | & (E) \end{array}$$

Figure 2. Grammar for experiments in a medical problem domain.

The order in which production rules are applied induces a characteristic bias. Therefore we modified LAGRAMGE’s original beam search approach in several ways, following Selman’s GSAT algorithm (Selman et al., 1992).

Randomization: Instead of choosing productions for the refinement of a parse tree node and the termination of the subtrees by the predefined sequential order, these are selected at random following the random restart hillclimbing algorithm GSAT, where a randomly generated truth assignment for a set of propositional clauses is set in a first step of the algorithm. Here, this helps to avoid local minima.

Lookahead: The best refinement of the current parse tree is selected as follows: Randomly choose a node in the current parse tree and refine it like described before. Iteratively traverse the resulting parse tree in preorder sequence until the result can not be improved anymore; in each iteration step apply each production applicable for the actually chosen node, randomly refine the corresponding subtree and evaluate the heuristic function. This forces the algorithm to try to improve each randomly refined successor hypothesis, even if it could not be improved by the first random refinement step. Quoting GSAT, this is similar to the step of flipping propositional variables. Again this helps to avoid local minima.

Momentum: The previously described iteration step can be interpreted as an unconstrained momentum, as it keeps the lookahead running through local minima as long as the actual hypothesis is improved for a minimal $\varepsilon = 0.0$. In general, this results in overfitting the data. To avoid this, a minimum improvement of $\varepsilon \geq$

0.0 is required for the preorder loop.

4. Experimental Evaluation

In the following we describe the experiments designed to show the competitiveness to the original algorithm, cf. Sec 4.1, the independence of the algorithm to input sequences, cf. Sec 4.2 and the applicability to real-world datasets, cf. Sec 4.3.

If applicable, we report on the dependent variable v_{dep} considered, the beam width w_{beam} used, the number of evaluated parse trees t_{eval} , the root mean squared error $rmse$, the random seed setting $seed$, the permutation of the order among the production rules $permute$, whether the original model was discovered $found?$, and the percentage of times the EOM was discovered $\%positive$.

4.1. Equation Discovery Benchmarks

To show the competitiveness of the approach to the original system, several experiments on standard problems from the field of equation discovery were conducted: 'Optimal Phytoplankton Growth', 'Aquatic Ecosystem', 'Springs and Masses', 'Tow Poles on Card', and 'Piecewise Linear Circuit'. Representative for the general results we present those on 'Piecewise Linear Circuit', as this domain has the highest number of deriveable parse trees, namely 18802. The piecewise linear circuit can be described by three equations, describing the first derivative of an x , y and z parameter:

$$\begin{aligned}x' &= \sigma(q - x), \\y' &= \mu_1(r - z) - qy, \\z' &= \mu_2x - bz.\end{aligned}$$

The dataset consists of 10 simulations, each for randomly chosen initial states. Overall, there are 1000 samples. The beam width was set to 1, 5 and 25. The results are summarized in Table 1.

As can be seen, competitive estimates for x' , y' and z' are produced by the novel system. Indeed, the algorithm considered more parse trees, i.e., the running time was higher. On the other hand, the memory consumption was lower, as a beam size of 1 is already sufficient in contrast to the beam size of 50 for the original approach.

4.2. Equation of Motion: Simulated Data

To show that the performance of the novel system is independent of the input order on both the productions of the grammar and the data and furthermore to

Table 1. Results for piecewise linear circuit, modeled by original (*Top*) and nested beam search (*Bottom*).

v_{dep}	w_{beam}	t_{eval}	RMSE	FOUND?
ORIGINAL SEARCH				
x'	1	8	3.099	×
y'	1	8	4.217	×
z'	1	28	0.684	×
x'	45	20	0.005	✓
y'	45	23	0.306	✓
z'	45	22	0.285	✓
x'	50	20	0.005	✓
y'	50	21	0.237	✓
z'	50	22	0.285	✓
NOVEL SEARCH				
x'	1	2588	0.005	✓
y'	1	3151	0.237	✓
z'	1	3219	0.285	✓

investigate the robustness concerning the randomization aspect, we conducted experiments in the controlled setting of simulated data in a medical problem domain. The aim was to rediscover the EOM from the simulated dataset.

We considered 6 permutations of the independent variables volume V , flow F and PEEP P . The 6 permutations were combined with 2 different orders of the production rules in the grammar, consisting of an initial order and its reverse. To investigate the robustness concerning randomization, we considered for each of these combinations 5 random seeds. Thus, we had 12 experimental runs for the original search algorithm and 60 for the novel approach.

To increase the complexity of the search problem, simulations for 10 different PEEP offsets were run, the resistance was simulated to be different for inspiration and expiration and the compliance was simulated dynamically with its maximum value set to 50.

A universal grammar for expressions using the $+$, $-$, $*$ and \div operators was used, cf. Fig 2. The range for the constants was set to $[0, 70]$, with 25 as start value for the fitting procedure. A depth of 5 for the parse tree would have been sufficient to produce the equation of motion. For the experiments, however, the maximum depth was set to 6, which allows to derive $1.5 * 10^7$ parse trees.

According to the nonlinear dependencies in the EOM, i.e. $Flow * const + Vol/const$ we say that a discovered model (i.e. one of the beam elements) describes the EOM if the two first two addends (or equivalents)

Table 2. Means of results for 12 production rule and variable input sequences; *Top*: Nested beam search for 5 random seed settings. *Bottom*: Original beam search.

SEED	RMSE	% POSITIVE
NOVEL SEARCH		
1	3.11 ± 0.73	83.3
2	3.05 ± 0.35	75.0
3	3.63 ± 1.00	41.7
4	3.19 ± 0.71	75.0
5	3.07 ± 0.73	75.0
MEAN	3.21 ± 0.70	70.0
ORIGINAL SEARCH		
-	4.83 ± 1.85	33.3

appeared in the model.

The results are summarized in Table 2. To avoid memory overflow, we had to restrict the beam width to 25 when running the original system, which found the EOM in 33% of the 12 runs. In contrast, the new approach discovered the EOM in about 70% of the runs. This is significantly better. Moreover, there was almost no effect by the random seed.

4.3. Equation of Motion: Real-World Data

To show the applicability of the novel approach to real-world datasets, we conducted experiments on data recorded from 4 patients suffering from the ARDS (Adult Respiratory Distress Syndrome), which indicates a highly damaged lung. Each patient dataset consisted of 12 to 14 measurements. Between each measurement the PEEP was increased by 2.0 mbar, starting from 0.0 mbar until a maximum pressure of 45.0 mbar was reached. The sample frequency was set to 8/sec. This form of measuring is called the PEEP-wave manoeuvre (Putensen et al., 1989).

The setup was the same as for the simulated dataset. We performed 12 runs for one random seed setting with a beam width of 1. The results are summarized in Table 3. The EOM was detected in 91.7% of the runs.

5. Conclusion

We presented an equation discovery system based on LAGRAMGE. By modification of the system, we used a randomized, hill-climbing heuristic search among the refinements.

In an extensive set of experiments we could show, that

Table 3. Results for real-world datasets with 12 input sequence permutations.

PERMUT	t_{eval}	RMSE	FOUND?
1	3084	4.80	✓
2	3247	4.80	✓
3	2035	7.36	×
4	2891	4.80	✓
5	2415	4.80	✓
6	2640	4.80	✓
7	3809	4.80	✓
8	2548	4.99	✓
9	4087	4.80	✓
10	6415	4.80	✓
11	2671	4.80	✓
12	2756	4.80	✓
MEAN:	3216.5	5.03	91.7%

this novel approach is indeed competitive to the standard system concerning quality, is sound towards an input induced bias and - most important - is applicable in real-world datasets.

To the best of our knowledge, this is the first application of an equation discovery technique to measured respiratory data from intensive care medicine. In this field we find an application potential for detecting and modelling regularities in pulmonary physiology.

Acknowledgement

This work was supported by the Deutsche Forschungsgesellschaft (DFG), Grant No. 561/6-1.

References

- DeRaedt, L. (Ed.). (1996). *Declarative bias in ilp*. Amsterdam: IOS Press.
- Putensen, C., Baum, M., Koller, W., & Putz, G. (1989). Peep-welle: Ein automatisiertes verfahren zur bettseitigen bestimmung der volumen/druck beziehung der lunge beatmeter patienten. *Anaesthesist*, 38, 214–219.
- Selman, B., Levesque, H. J., & Mitchell, D. (1992). A new method for solving hard satisfiability problems. *Proceedings of the Tenth National Conference on Artificial Intelligence* (pp. 440–446). Menlo Park, CA: AAAI Press.
- Todorowski, L., & Džeroski, S. (1997). Declarative bias in equation discovery. *Proceedings of Fourteenth International Conference on Machine Learning* (pp. 376–384). San Francisco: Morgan Kaufmann.