LAGRAMGE - an equation discovery system: User's manual

1 Introduction to LAGRAMGE

LAGRAMGE [1, 2, 3] is an equation discovery system capable of discovering differential as well as ordinary algebraic equations involving more than two variables from observational data only. It allows the user to specify the form of the equations with a context free grammar. The grammar can use the usual mathematical operators defined in the C programming language, as well as additional functions defined by the grammar at hand. Domain specific knowledge is used to specify the grammar, focusing the equation discovery process on equations with structure that is acceptable and comprehensible within the domain of use. For additional information consult [1, 2, 3].

2 Installation

LAGRAMGE is implemented in the C programming language using the Gnu `gcc` compiler. It was developed on a Hewlett Packard workstation working under the HP-UX operating system. It was successfully compiled and run on Hewlett Packard (HP-UX 10.01) and PC (under the Linux RedHat 6.1 operating system) workstations.

Once you have retrieved the distribution file `lagramge-www.tar.Z`, you have to extract the files with the following command:

```
% uncompress lagramge-www.tar.Z ; tar xf lagramge-www.tar
```

Before building the executables you should edit the following line in the file `Makefile` in the directory `Lagramge/src`:

```
LYCFLAGS = -g -O -DHOME_DIR="/home/ljupco/ml/Lagramge/src"
```

to reflect the path, where the LAGRAMGE is installed on your workstation.

To build the executables for running LAGRAMGE, type `make all` in the directory `Lagramge/src`. Later, after the software starts working, you can remove the compiled code (`.o` files) and other intermediate files by typing `make clean` in the same directory.

The directory `Lagramge/doc` contains postscript versions of some of the papers referenced in this manual and the manual itself. Several examples of experiments in different domains are included within this distribution in the subdirectories of the source directory (`Lagramge/src`).

3 Using LAGRAMGE

The process of equation discovery with LAGRAMGE is divided in two phases. In the first phase a context free grammar is compiled for use in LAGRAMGE with the command:

% `gc` *grammmar-file-name*

where *grammar-file-name* is the name of the file that contains the context free grammar along with the definitions of the background knowledge functions used. The format of the file is described in Appendix A. The second, main phase starts LAGRAMGE on the domain at hand:

```
% g -g grammar-name -v variable-name [-h time-step] [-t time-variable]
  [-c config-file-name] [-C] domain-file-name [domain-file-name(s)]
```

The command accepts the following switches and parameters:

`-g` *grammar-name* - determines the grammar to be used in LAGRAMGE. The grammar should be compiled, as described before.

`-v` *variable-name* - determines the dependent variable from the domain. The use of this switch is obligatory, because one of the system variables has to be distinguished as dependent.

`-h` *time-step* - provides the time step between two successive measurements for discovering differential equations. If this and the following switch are omitted, ordinary equations are sought by LAGRAMGE.

`-t` *time-variable* - in case when measurements are taken at non-equidistant time points, the time variable can be specified.

`-c` *config-file-name* - provides the name of the file with LAGRAMGE's parameters. If this switch is omitted the default configuration file `default.cfg` is used:

```
search_strategy: beam
fitting_algorithm: amoeba
heuristic_function: sse
stopping_criterion: none
max_derivation_tree_depth: 5
beam_width: 25
```

By default, LAGRAMGE uses beam search strategy with a beam width of 25, the downhill simplex method for constant parameters' fitting, the *SSE* heuristic function, the stopping criterion $\mathcal{C} = \textbf{false}$ and maximal height of derivation trees $h_{max} = 5$. For more information on the format of the configuration parameters' file see Appendix A.

`-C` - the equations are generated in some predefined canonical (lexical) order, so the redundant equations resulting from the commutativity of `*` and `+` operators are omitted. The code behind this option was added to LAGRAMGE very recently and is still in a preliminary state. Use this option with caution.

domain-file-name(s) - one or more file names. Each file contains a table of measurements of the domain variables at successive time points.

The output of LAGRAMGE is the list of best equations found, along with their heuristic function values, the number of considered derivation trees and the elapsed CPU time.

4 Contact address

Ljupčo Todorovski
Department of Intelligent Systems
Jožef Stefan Institute
Jamova 39
1000 Ljubljana
Slovenia
e-mail: `Ljupco.Todorovski@ijs.si`
home-page: `http://www-ai.ijs.si/~ljupco/`

The most recent version and information about the (further) development of LAGRAMGE is available at `http://www-ai.ijs.si/~ljupco/ed/`.

References

[1] Todorovski, L. and Džeroski, S. (1997) Declarative bias in equation discovery. In *Machine learning. Proceedings of the 14th international conference (ICML'97)*, pages 376–384. Morgan Kaufmann publishers, San Francisco, CA.

[2] Todorovski, L. (1998) *Declarative bias in equation discovery.* M.Sc. Thesis. Faculty of Computer and Information Science, Ljubljana, Slovenia.

[3] Todorovski, L., Džeroski, S. and Kompare, B. (1998) Modelling and prediction of phytoplankton growth with equation discovery. *Ecological Modelling* 113: 71–81.

A Lagramge input files

A.1 Grammar files

$$
\begin{array}{lll}
\textit{GrammarFile} & \rightarrow & \textit{FunctionDefinitions Productions} \\
\\
\textit{FunctionDefinitions} & \rightarrow & \text{“\%\{“ } \mathbf{function-definitions-in-C} \text{ “\%\}“} \mid \epsilon \\
\\
\textit{Productions} & \rightarrow & \textit{Production Productions} \mid \textit{Production} \\
\\
\textit{Production} & \rightarrow & \textit{LeftSide} \text{ “}\rightarrow\text{“ } \textit{RightSide} \text{“;“} \\
\textit{LeftSide} & \rightarrow & \textit{Nonterminal} \\
\textit{RightSide} & \rightarrow & \textit{Symbols} \text{ “|“ } \textit{RightSide} \mid \textit{Symbols} \\
\\
\textit{Symbols} & \rightarrow & \textit{Symbol Symbols} \mid \textit{Symbol} \\
\textit{Symbol} & \rightarrow & \textit{NonTerminal} \mid \textit{Terminal} \\
\\
\textit{NonTerminal} & \rightarrow & \textit{CapitalLetter} \mid \textit{String} \\
\textit{Terminal} & \rightarrow & \textit{String} \\
\\
\textit{String} & \rightarrow & \textit{String AlphaNum} \mid \epsilon \\
\\
\textit{AlphaNum} & \rightarrow & \textit{Letter} \mid \textit{Digit} \mid \textit{Misc} \\
\textit{Letter} & \rightarrow & \text{“}\mathbf{a}\text{“} \mid \text{“}\mathbf{A}\text{“} \mid \text{“}\mathbf{b}\text{“} \mid \text{“}\mathbf{B}\text{“} \mid \ldots \mid \text{“}\mathbf{z}\text{“} \mid \text{“}\mathbf{Z}\text{“} \\
\textit{CapitalLetter} & \rightarrow & \text{“}\mathbf{A}\text{“} \mid \text{“}\mathbf{B}\text{“} \mid \ldots \mid \text{“}\mathbf{Z}\text{“} \\
\textit{Digit} & \rightarrow & \text{“}\mathbf{0}\text{“} \mid \text{“}\mathbf{1}\text{“} \mid \ldots \mid \text{“}\mathbf{9}\text{“} \\
\textit{Misc} & \rightarrow & \text{“}+\text{“} \mid \text{“}-\text{“} \mid \ldots
\end{array}
$$

A.2 Configuration files

$$
\begin{array}{lll}
\textit{ConfigurationFile} & \rightarrow & \textit{ParameterSettings} \\
\textit{ParameterSettings} & \rightarrow & \textit{ParameterSetting ParameterSetting} \mid \textit{ParameterSetting} \\
\textit{ParameterSetting} & \rightarrow & \mathbf{parameter-name} \text{ “:“ } \mathbf{parameter-value}
\end{array}
$$

Parameter name	Possible values
`search_strategy`	`beam` `exhaustive`
`fitting_algorithm`	`amoeba` `lm`
`heuristic_function`	`sse` `mdl`

`stopping_criterion`	`none`
	`error_treshold`
	`time_limit`

`max_derivation_tree_depth`	positive integer number
`beam_width`	positive integer number
`error_treshold`	positive real number
`time_limit`	positive real number

A.3 Domain files

$$
\begin{array}{lcl}
\mathit{DomainFile} & \rightarrow & \mathit{VarNames}\ \mathit{Measurements} \\
\\
\mathit{VarNames} & \rightarrow & \mathit{VarName}\ \mathit{VarNames} \mid \mathit{VarName} \\
\mathit{VarName} & \rightarrow & \mathit{Letter}\ \mathit{String} \\
\\
\mathit{Measurements} & \rightarrow & \mathit{Measurement}\ \text{``;``}\ \mathit{Measurements} \mid \mathit{Measurement}\ \text{``;``} \\
\mathit{Measurement} & \rightarrow & \mathit{Measurement}\ \mathit{Number} \mid \mathit{Number} \\
\\
\mathit{String} & \rightarrow & \mathit{String}\ \mathit{AlphaNum} \mid \epsilon \\
\\
\mathit{Number} & \rightarrow & \mathit{Sign}\ \mathit{Digits} \mid \mathit{Sign}\ \mathit{Digits}\ \text{``.``}\ \mathit{Digits} \\
 & \mid & \mathit{Sign}\ \mathit{Digits}\ \text{``.``}\ \mathit{Digits}\ \text{``}\mathbf{E}\text{``}\ \mathit{Sign}\ \mathit{Digits} \\
\mathit{Digits} & \rightarrow & \mathit{Digits}\ \mathit{Digit} \mid \mathit{Digit} \\
\\
\mathit{AlphaNum} & \rightarrow & \mathit{Letter} \mid \mathit{Digit} \mid \text{``\_``} \\
\mathit{Letter} & \rightarrow & \text{``}\mathbf{a}\text{``} \mid \text{``}\mathbf{A}\text{``} \mid \text{``}\mathbf{b}\text{``} \mid \text{``}\mathbf{B}\text{``} \mid \ldots \mid \text{``}\mathbf{z}\text{``} \mid \text{``}\mathbf{Z}\text{``} \\
\mathit{Digit} & \rightarrow & \text{``}\mathbf{0}\text{``} \mid \text{``}\mathbf{1}\text{``} \mid \ldots \mid \text{``}\mathbf{9}\text{``} \\
\mathit{Sign} & \rightarrow & \text{``}+\text{``} \mid \text{``}-\text{``} \mid \epsilon
\end{array}
$$