John's PhD Thesis

A subtitle of software implementations

John Zobolas

2019-08-07

Contents

List of Tables

List of Figures

Chapter 1

Preface

What is this thesis about?

1.1 Introduction

How man became interested in modeling...

Chapter 2

PhD work finished

Just a summary of all the work (the "real" results - investigations/"secondary" things are not included - that I have achieved in my PhD until now (to write in bookdown soon).

2.1 Pipeline

- First refactoring, bug fixing, source code documentation on Gitsbe, Drabme, druglogics2 (druglogics-synergy) modules and most important parallel simulations enabled (my first 2-3 months of my PhD)
- Second refactoring, transport modules to maven packaging, added support for many features (Ongoing work - see dev_plan_doc). Most important things achieved:
  - Added tests to modules Gitsbe, Drabme usign JUnit5, mockito, assertJ libraries
  - druglogics-roc-generator R shiny app
  - Export support using BioLQM: the initial model + best generation models can now be exported through configuration options to GINML, SBML-Qual and BoolNet community formats

2.2 VSM

- VSM Dictionary Bioportal module

2.3 PSICQUIC

- miTab 2.8 support added to psicquic web service
  - See psicquic doc here
  - miTab 2.8/causalTab paper
- Update JAMI library to support miTab 2.8 - results of the Bio-Hackathon 2018, Paris and the Marseille GREEKC hackathon event.

2.4 Others

- Java Client for RSAT tool fetch-sequences

Chapter 3

PhD Plans

Plans are currently in terms of technologies. This will change.[1]

3.1 Pipeline

- Ensemble Model biomarker analysis (part of this work is for automated pipeline paper). Roadmap:
  - Make R package with general useful functions
  - Make R package for biomarker analysis
  - Make bookdown document for previous atopo-based analysis
  - Redo the analysis on Cascade topology and for specific drug combinations (2)
  - Feature importance/biomarker selection using ML methods (compare with what you got with your own method, submit to the ML course to get the credits)
  - Submit R packages to CRAN
  - Small publication of the R package and ML methods perhaps?
- Work on the pipeline modules (see the dev_plan_doc). Most important:
  - Full Testing (Junit 5)
  - BioLQM support: stable state calculation, trap spaces

[1] Perhaps

  - Do comparison between Aurelien's BioLQM stable state algorithm and BNReduction using M2

3.2 VSM

- Make ontologies for genes and proteins known databases (Ensembl, Uniprot)
- PubDictionaries
- Dictionary Merger/Combiner

3.3 PSICQUIC

- Connect data taken from psicquic to atopo (enxtend atopo module to be PSICQUIC-compatible - takes causality information and builds network of interactions with available configuration on how to do so).
- Help build PSICQUIC 2.0

3.4 Synergy

- Augment existing R package for calculating reference models to include Wim's generalized Bliss method and the mean synergy score by Simone Laderer! Goal is to test all the null reference models (Loewe, Bliss, ZIP, +2 new, others?) on read dose-response matrix datasets and see which is best at finding the synergies. R package to use:
  - SynergyFinder from Finland group code here
  - Also see this software: R package COMBIA
- Mathematical formulation of the volume-based synergy score general method to include (mine)?

Miscellaneous Stuff

Chapter 4

PhD ideas

4.1 Compare fixpoint tools

Compare different tools that calculate fixpoints for logical modeling.

Models used for testing could be of different types:

- self-contained
- varying the number of input nodes (1-n)
- small to large number of nodes
- small to large number of edges
- scale-free (boolnet generated) vs random (varying K connectivity)
- play with form of the boolean equations
- others ???

Other things that can be done:

- support BNReduction data format by Veliz-Cuba in BioLQM
- add support for calculating the fixpoints using the Colomoto docker (python interface)
- comparison between **BioLQM, Pint, MABOSS and BNReduction** could be done then in a Jupiter colomoto-enabled notebook!

Further extension/comparisons could be:

- Tamura, 2009 - Integer programming method
- Dubrova, 2009 - SAT-based

4.2 Use Logical modeling to predict single-drug data

Asmund project proposal: mechanistic drug response prediction analysis

- Automate drug target profile annotation from:
  - Klaeger publication 2017 Science
  - mrc ppu
  - Davis publication 2011 (nature biotechnology?)
- Omics data (rna, cnv etc)
  - COSMIC
  - CCLE
- Drug scren data
  - Single drug
    - COSMIC/GDSC
    - CCLE
  - Combo
    - O'Neil 2016 Molecular cancer therapeutics
    - FDA Holbeck 2017 cancer research publication

My idea is more like this:
Predict drug-response curves from drug combination datasets (GDSC, CCLE), using logical modeling for singaling network analysis or translation from logical to ODE modeling. Aslo try to predict drug combinations datasets (dose-response matrices?). Pretty much what is done in this paper with help from this one for converting boolean models to continuous.

4.3 Quantum logic formalism

Instead of logical modeling formalism, use the idea of (queantum) logical gates. The **core idea** makes sense: you don't know the state of a protein, but when you measure it, only then you really know what it is.
May also be worth to look at a game-theoritic approach to find attractors and such.

4.4 Harmony Search

Use this algorithm for optimizing the boolean equations for gitsbe?

4.5 Druglogics-Pipeline related

4.5.1 Train models to cell-specific proliferation

Concept is that random models predict a lot better than cell-specific ones: main directive is **proliferation**, not just fitting to a steady state pattern.

4.5.2 A bottom-up model building for drug prediction

Start with a model and some observed synergies. Build/train/produce models that predict the first observed synergy (using Harmony Search?), from them the next one, etc. You end up with many models that can predict all the observed synergies or you try to find out why that cannot happen for example (e.g. contrasting synergies?). Do the latest models' stable states or attractors correspond to activity of proteins from literature?

4.5.3 Simulate cancer resistance

For example, you have some models that predict some (observed) synergies or you just find some synergistic drug comibnations for these models or per model. Then, you modify these models in order to be resistant to these drugs, simulating thus the cancer rewiring process! Then, you apply (n+1) drug combinations to win over the resistance (and you do this procedure at more levels to suggest 3-way, 4-way drug combos and why there might be cancer models that can 'win' over these models and continue the proliferation). You end up with super cancer resistant models and methods to achieve them or reasons why this cannot happen at all.

Appendix A

Bookdown features

Testing stuff!
Nice link for reference on how to write a thesis with bookdown.

A.1 References

- Important to remember: Label can be: fig:foo, thm:foo
- Appendix won't have references (figures, tables, equations ain't gonna be there)
- Examples
  - VSM, Compare fixpoint tools
  - A link to Pipeline
  - Table stuff
  - A paper citation: Albert et al. (2008)
  - Same paper citation: (Albert et al. 2008)

A.2 Markdown examples

An epigram:

> "I thoroughly disapprove of duels. If a man should challenge me,

> I would take him kindly and forgivingly by the hand and lead him to a quiet place and kill him."
>
> — Mark Twain

$f = a$

$$f = a \tag{A.1}$$

See Equation (A.1).

Theorem A.1 (Pythagorean theorem)**.** *For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the other two sides, we have*

$$a^2 + b^2 = c^2$$

See Theorem A.1.

Definition A.1 (Pythagorean theorem)**.** For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the other two sides, we have

$$a^2 + b^2 = c^2$$

See Definition A.1.

A.3 Figures

A normal paragraph.

See Figure A.2!

A.4 Tables

See Table A.1

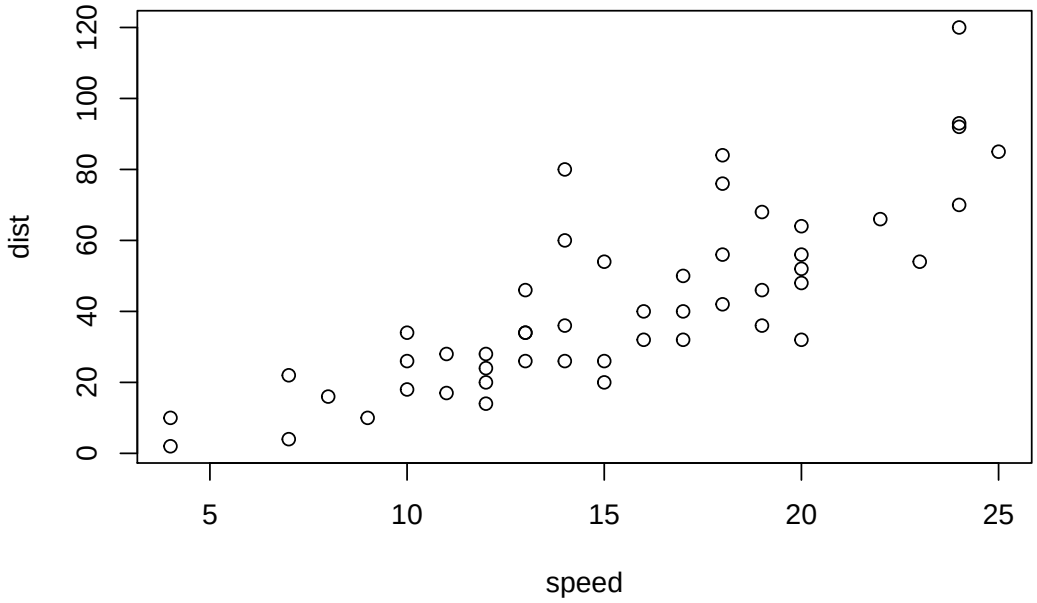

Figure A.1: A scatterplot of the data `cars` using **base** R graphics.

Figure A.2: A Big Question

Table A.1: A table of the first 5 rows of the mtcars data.

	mpg	cyl	disp	hp	drat	wt	qsec	vs
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0

```
DT::datatable(iris)
```

Show 10 entries Search:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa

Showing 1 to 10 of 150 entries Previous 1 2 3 4 5 ... 15 Next

Appendix B

Another appendix!!!

Software used, inspirations, etc.

References

Albert, István, Juilee Thakar, Song Li, Ranran Zhang, and Réka Albert. 2008. "Boolean network simulations for life scientists." *Source Code for Biology and Medicine* 3: 16. https://doi.org/10.1186/1751-0473-3-16.