A First-Order Axiomatization for Transition Learning with Rich Constraints

Jáchym Barvínek, Filip Železný

Czech Technical University in Prague

{barvijac,zelezny}@fel.cvut.cz

September 3, 2017

Problem overview

- Reminder - a Boolean network:
  - is essentially a function: $BN : Bool^n \rightarrow Bool^n$,
  - iterated from a state eventually reaches an attractor,
  - can be represented by n propositional formulas.

Problem overview

- Reminder - a Boolean network:
  - is essentially a function: $BN : Bool^n \to Bool^n$,
  - iterated from a state eventually reaches an attractor,
  - can be represented by n propositional formulas.
- Our task: Learn a Boolean network model given:
  - Observed data (List of $Bool^n \times Bool^n$)
  - Constraints on attractor behaviour.
  - Constraint on the type of formulas representing the BN.

Considerations for applications in bioinformatics

- Observed data – scarce.
- Attractor constraints – may be known.
  - e.g.: Cell eventually reaches homeostasis, dies etc.
- Formula constraints – reasonable assumptions normally hold in practical models.
  - e.g.: Both Mammalian Cell Cycle (n=9)[1] and Fanconi Anemia (n=28)[2] models can be represented by 4-term DNFs.

[1] Adrien Fauré et al. "Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle". In: *Bioinformatics* 22.14 (2006).

[2] Alfredo Rodríguez et al. "A Boolean network model of the FA/BRCA pathway". In: *Bioinformatics* 28.6 (2012).

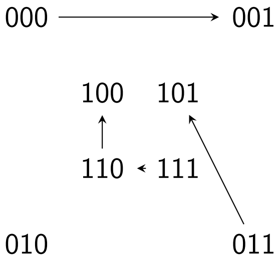
Why constraint the model?

- Theorem: To learn a unique, correct BN from data only, we need to know transitions from all possible states.[3]
  - No free lunch for BN learning.
- We show that adding the constraints reduces the need for data.

[3] D. Cheng, H. Qi, and Z. Li. *Analysis and Control of Boolean Networks: A Semi-tensor Product Approach*. Springer London, 2010.

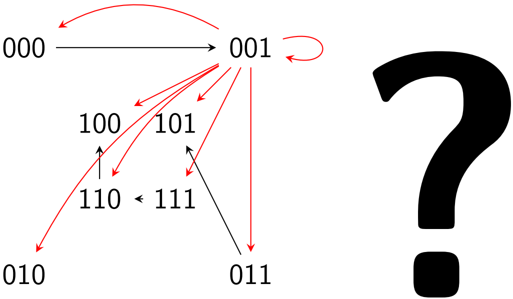
Example

Given observations



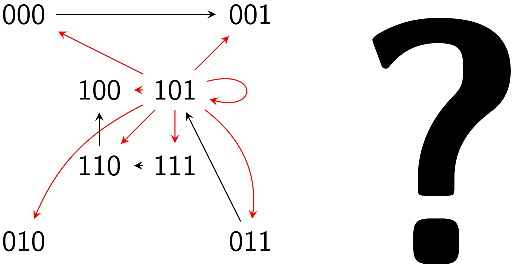
Example

2^n possible transitions from each unseen state. Which one to choose?



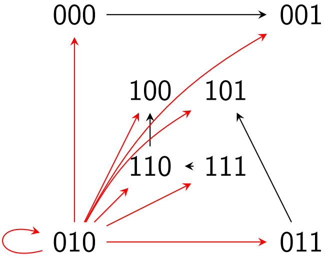
Example

2^n possible transitions from each unseen state. Which one to choose?



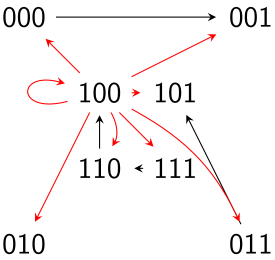
Example

2^n possible transitions from each unseen state. Which one to choose?



Example

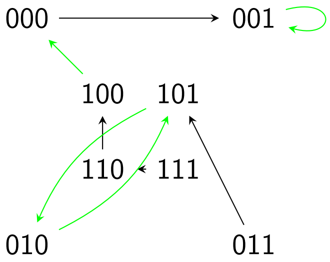
2^n possible transitions from each unseen state. Which one to choose?



Example

With constraints

- When we use the following constraints, there is a unique admissible model:
  - Exactly 2 attractors; one of period 2 and one fixed point.
  - Formula for each variable is a conjunction of literals.



$$p(t+1) = q(t)$$
$$q(t+1) = p(t) \wedge r(t)$$
$$r(t+1) = \neg p(t)$$

Identifying the solutions

A provisional algorithm

- Transform the constraints and data to axioms of a FOL theory.

Identifying the solutions

A provisional algorithm

- Transform the constraints and data to axioms of a FOL theory.
- Use a model-checking technique to find its model(s).[4]

[4] Koen Claessen and Niklas Sörensson. "New techniques that improve MACE-style finite model finding". In: *Proceedings of the CADE-19 Workshop: Model Computation-Principles, Algorithms, Applications*. 2003.

Identifying the solutions

A provisional algorithm

- Transform the constraints and data to axioms of a FOL theory.
- Use a model-checking technique to find its model(s).[4]
- Transition formulas can be read directly from the model.

[4] Koen Claessen and Niklas Sörensson. "New techniques that improve MACE-style finite model finding". In: *Proceedings of the CADE-19 Workshop: Model Computation-Principles, Algorithms, Applications*. 2003.

Identifying the solutions

A possible formalization as a theory in FOL with equality

- Axiom for transition observation (example 000 $\rightarrow$ 001):
  $p(f,f,f) = f \wedge q(f,f,f) = f \wedge r(f,f,f) = t$
- We must explicitly state that: $\neg(f = t)$

Identifying the solutions

A possible formalization as a theory in FOL with equality

- Axiom for transition observation (example 000 → 001):
  $p(f,f,f) = f \wedge q(f,f,f) = f \wedge r(f,f,f) = t$
- We must explicitly state that: $\neg(f = t)$
- Define semantics of propositional conjunction:

$$\forall P_p, P_q, P_r, N_p, N_q, N_r, I_p, I_q, I_r.\ conj(P_p, P_q, P_r, N_p, N_q, N_r, I_p, I_q, I_r) = t \leftrightarrow \bigwedge_{i \in \{p,q,r\}} (P_i = t \rightarrow I_i = t) \wedge (N_i = t \rightarrow I_i = f)$$

- Constraint expressivity of p to conjuncion:
  $\exists P_p, P_q, P_r, N_p, N_q, N_r. \forall I_p, I_q, I_r.\ p(I_p, I_q, I_r) = conj(P_p, P_q, P_r, N_p, N_q, N_r, I_p, I_q, I_r)$
  - Formulas can be read from Skolem constants in a model.

Identifying the solutions

A possible formalization as a theory in FOL with equality

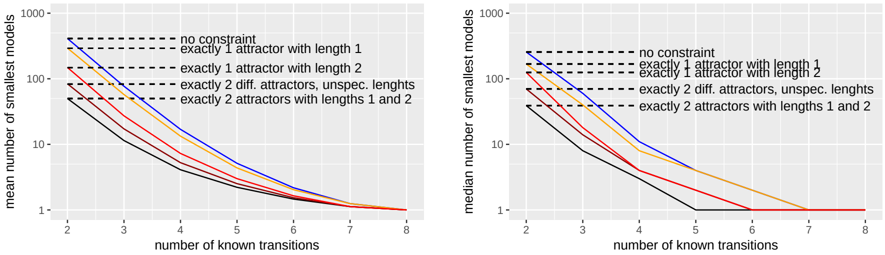
- Axiom for transition observation (example 000 → 001):
  $p(f,f,f) = f \wedge q(f,f,f) = f \wedge r(f,f,f) = t$
- We must explicitly state that: $\neg(f = t)$
- Define semantics of propositional conjunction:

$$\forall P_p, P_q, P_r, N_p, N_q, N_r, I_p, I_q, I_r.\ conj(P_p, P_q, P_r, N_p, N_q, N_r, I_p, I_q, I_r) = t \leftrightarrow \bigwedge_{i \in \{p,q,r\}} (P_i = t \rightarrow I_i = t) \wedge (N_i = t \rightarrow I_i = f)$$

- Constraint expressivity of p to conjuncion:
  $\exists P_p, P_q, P_r, N_p, N_q, N_r. \forall I_p, I_q, I_r.\ p(I_p, I_q, I_r) = conj(P_p, P_q, P_r, N_p, N_q, N_r, I_p, I_q, I_r)$
  - Formulas can be read from Skolem constants in a model.
- Constraint "There is a fixed-point attractor":
  $\exists I_p, I_q, I_r.\ p(I_p, I_q, I_r) = I_p \wedge q(I_p, I_q, I_r) = I_q \wedge r(I_p, I_q, I_r) = I_r$
- To require higher order attractors, iterate the transition function or use a reachability predicate.
- To restrict other attractors, use also uniqueness quantification.

Experimental results

On the above model with conjunction expressivity constraint



- Note: Without the conjunction constraint, the number of admissible models would be far higer: $8^{8-|data|}$.
- With the strongest constraints, $\frac{5}{8} = 62.5\%$ data was usually enough to unambiguously identify the single correct model.
- Conclusion: Knowledge about attractors can help us reduce the need for data in BN learning.

Presentation bibliography

D. Cheng, H. Qi, and Z. Li. *Analysis and Control of Boolean Networks: A Semi-tensor Product Approach*. Springer London, 2010.

Koen Claessen and Niklas Sörensson. "New techniques that improve MACE-style finite model finding". In: *Proceedings of the CADE-19 Workshop: Model Computation-Principles, Algorithms, Applications*. 2003.

Adrien Fauré et al. "Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle". In: *Bioinformatics* 22.14 (2006).

Alfredo Rodríguez et al. "A Boolean network model of the FA/BRCA pathway". In: *Bioinformatics* 28.6 (2012).

Thank you for attention!