Parameterized Complexity Analysis

A New and Powerful Tool
in the Cognitive Scientist's Toolbox

Johan Kwisthout, Radboud University Nijmegen /
Leiden Institute for Advanced Computer Science

Take home message

are to

as

is to

Outline

- Bayesian models of cognition
- Why computational complexity matters
- “HELP! My model's intractable! What now?”
- Analyzing sources of complexity
- Less is better – constrain your models
- Explaining and predicting behavior
- Take home message: use your toolbox!

“If I have seen further than others, it is by standing on the shoulders of giants” – Sir Isaac Newton

Todd Wareham

Joint work with Mark Blokpoel (Radboud University), Todd Wareham (University of Newfoundland), Iris van Rooij (Radboud University)

Cognitive modeling

- Computational models of cognition
  - Understand how the mind works
  - Predict human behavior (HCI)
  - Artificial intelligence / robotics
- Marr's hierarchy of analysis:
  - Computational level (what)
  - Algorithmic level (how)
  - Implementational level (realisation)
- Study, experiment, simulate, predict

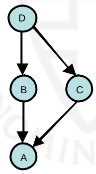
Bayesian models of cognition

- Many computational models nowadays are based on **Bayesian abduction**
- Bayesian abduction = inferring the *most probable explanation* of a set of observed phenomena
  - What are this person's goals given what I observe as his actions?
  - What does she want to communicate here?
  - What is the object that is partially occluded in my line of vision?

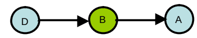
Bayesian networks

- **Bayesian network**: models a set of stochastic variables and the independency relations among them
- Directed acyclic graph with nodes and arrows; probability distribution for every node

- A (directly) depends on B and C
- The probability distribution of A is conditioned on the values of B and C
- B and C (directly) depend on D
- Other dependencies between variables depend on observations in the network

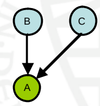


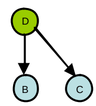
Conditional dependencies



- If **B** is observed, **A** is independent of **D** as there is no direct link: **A**'s probability distribution is governed by **B** only

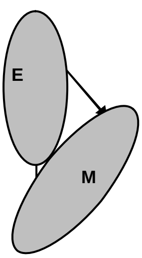
- If **A** is observed, **B** and **C** become dependent on each other as information on **B** 'explains away' **C** vice versa





- If **D** is observed, **B** and **C** become independent from each other as **D** is a common cause of **B** and **C**

Bayesian abduction in a nutshell

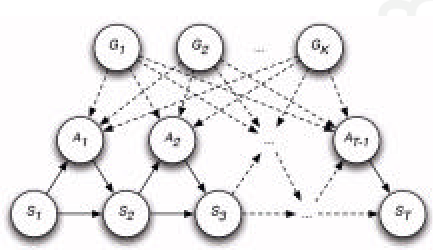


- **Input**: A Bayesian network partitioned in two sets **M** and **E**, and an observation **e** for the variables in **E**
- **Output**: The most probable joint value assignment **m** to **M** with **E** = **e**, or $argmax_m$ Pr(**M** = **m**, **E** = **e**)
- Bayesian abduction or *Most Probable Explanation* happens to be intractable (*NP*-hard) in general

The Bayesian Inverse Planning model

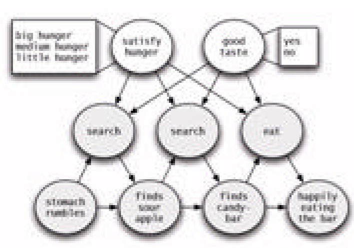
- Baker et al. (2009): goal recognition = inverse planning
  - We can hypothesize about someone's goals from the observed states of the world and her actions in that world
  - Example:
  - It's 7 AM. You observe me leaving my house, apparently on my way to the railway station. After walking 100 meters, I stand still, search my pockets, then run back home, only to return back in a minute and quickly walk in the direction of the station again.
  - Explain my goals at the moment of leaving the house for the first time, given what you have observed (actions / states)

The Bayesian Inverse Planning model



Baker, C.L., Saxe, R. and Tenenbaum, J.B. (2009). Action understanding as inverse planning. *Cognition*, 113 (3), 329-349.

An example: the hungry boy



The curse of intractability

- Bayesian abduction happens to be *NP*-hard (Shimony, 1994 - reduction from Vertex Cover; Kwisthout, 2011- reduction from SAT) even for binary variables
- The BIP-model (which is a special case of Bayesian abduction) is in general also *NP*-hard (Blokpoel, Kwisthout, van der Weide, en van Rooij, 2010); also even for binary variables

Why does it matter?

- Recall: NP-hard means: no polynomial worst-case algorithm possible unless P=NP

N	**10**	**50**	**100**	**300**	**1000**
$\log_2 N$	3	5	6	8	9
$5 \cdot N$	50	250	500	1,500	5,000
$N \cdot \log_2 N$	33	282	665	2,469	9,966
N^2	100	250	10,000	90,000	1,000,000
N^3	1,000	125,000	1,000,000	$2.7 \cdot 10^7$	$1.0 \cdot 10^9$
2^N	1,024	$1.1 \cdot 10^{15}$	$1.3 \cdot 10^{30}$	$2.0 \cdot 10^{90}$	$1.0 \cdot 10^{301}$
$N!$	3,628,800	$3.0 \cdot 10^{64}$	$9.3 \cdot 10^{157}$	$3.1 \cdot 10^{614}$	$4.0 \cdot 10^{2567}$

- Not polynomial = intractable for all but very small inputs

So what?

- *NP*-hardness means that in *general* there cannot exist a polynomial-time algorithm for solving *arbitrary* instances of Bayesian abduction (or BIP)
- Consequently, there are instances ***that are valid model instances*** that cannot be computed in polynomial time – and thus, the validity of the computational model of the cognitive task is at stake
- "Hey, my model does not encode SAT formulas and the like, that is not a real world problem!"

Begging the question

- ***Socrates:*** "Your model assumes *NP*-hard computations!"
- ***Cebes:*** "*NP*-hardness doesn't say anything. Of course there are instances that my model doesn't compute in polynomial time. But these are unrealistic instances. My model does well on reasonable instances"
- ***Socrates:*** "Fine. Which are those reasonable instances?"
- ***Cebes:*** "Well, those instances that my model computes in polynomial time, of course!"

Now what?

- As Socrates pointed out, we cannot just ignore NP-hardness – this issue needs to be addressed (remember our goal was to **understand** how the mind works!)
- **"Now what?" – three ways of dealing with intractability**
  - The *doomsday* approach
  - The *hand-waving* approach
  - The *analytical* or *rational* approach

The doomsday approach

Bayesian abduction is *NP*-hard

Bayesian models are no good models of the brain

- The **pessimists** throw away the baby with the bath-water: because Bayesian abduction is *NP*-hard, that doesn't rule out that *many* instances of abduction problems can be solved tractably

The hand-waving approach

Bayesian abduction is *NP*-hard

OK, fine; we'll just assume that the mind *approximates* Bayesian abduction then…

- The **optimists** try to solve the problem by asserting that approximation, satisficing, and using heuristics will be sufficient to overcome intractability. However, approximating Bayesian abduction and satisficing is *NP*-hard as well (Kwisthout, 2011; Kwisthout, van Rooij and Wareham, 2011)!

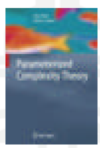
The analytical approach

Bayesian abduction is NP-hard

How can we *constrain* our model such that abduction becomes tractable again?

- The **realists** see the strength of Bayesian models but acknowledge that they are too broad and need to be constrained in order to overcome intractability. They will look for ***problem parameters*** that – when bounded – render the problem tractable

Parameterized complexity theory



- Even when a problem ? is NP-hard in general, it may be the case that there exist particular problem parameters, such that the problem can be solved tractably if the parameter is low.
- Formally, a problem with input size n may have parameters $k_1, k_2, \ldots, k_n$ and an algorithm solving the problem in time $O(f(k_1,k_2,\ldots,k_n) \cdot n^c)$ for an arbitrary computable function f and a constant c
- Hence, when $k_1, k_2, \ldots, k_n$ are small (enough), the running time of the algorithm is dominated by the $O(n^c)$ factor

Parameterized complexity analysis of MPE

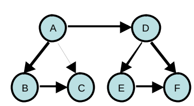
- Parameters that – when small – render Bayesian abduction tractable:
  - One minus the probability of the most probable explanation (i.e., when the probability of the MPE is high)
  - The *treewidth* of the network *and* the number of possible values per variable (both need to be small)
- Parameters that – even when small – do *not* render Bayesian abduction tractable:
  - The degree of the network, i.e., the number of incoming/outgoing arcs
  - The number of possible values per variable alone
- Other parameters are yet undecided
  - Treewidth alone, range of the probability distribution, …

Treewidth

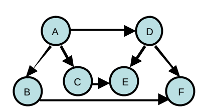
- The *treewidth* of a graph is a theoretical concept that loosely correlates to a measure on the *localness of the connections* in the graph
- If connections tend to be clustered in small sub-networks, with few connections between them, treewidth often is low
- If connections are scattered all over the place, treewidth may be high
- Many NP-hard graph problems are tractable when the treewidth of the graph is small

Hans Bodlaender

Two examples



- Two distinct clusters with only one connection
- Treewidth happens to be 2



- No distinct clusters, connections all over the place
- Treewidth happens to be 4
- Intuitive idea: computations are easier when they are localized

~~Begging~~ Answering the question

- ***Socrates:*** "Your model assumes NP-hard computation!"
- ***Cebes:*** "NP-hardness doesn't say anything. Of course there are instances that my model doesn't compute in polynomial time. But these are unrealistic instances. My model does well on reasonable instances"
- ***Socrates:*** "Fine. Which are those reasonable instances?"
- ***Cebes:*** "Well, those instances in which parameters k_1, k_2, ..., k_n are small!"

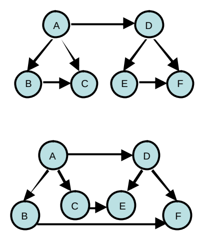
Socrates revisited

- ***Socrates:*** "Which are those reasonable instances?"
- ***Cebes:*** "Well, those instances in which parameters k_1, k_2, ..., k_n are small"
- ***Socrates***: "Ah, but are they small in practice?"
- ***Cebes:*** "I don't know, but let's ask a cognitive scientist to see whether she thinks that it is plausible that k_1, k_2, ..., k_n are typically small in cases where humans perform the cognitive task easily"

Socrates revisited

- ***Cebes:*** "Dear cognitive scientist, do you think that k_1, k_2, ..., k_n are typically small in cases where humans perform these cognitive tasks easily?"
- ***Cognitive Scientist:*** "Hmm, well, I'm pretty sure that k_1, k_2, ..., k_{n-1} are, but I'm not sure about k_n really..."
- ***Socrates***: "So, Cebes, how could you verify whether k_n is indeed small in practice and thus that your model is a good description of reality?
- ***Cebes:*** "Well, er, ... let's design an experimental setting with two comparable scenarios in which a cognitive task is measured, *that differs only in* k_n, and measure reaction times and error rates. If my model is right, performance will lower significantly when k_n increases!"

Possible setup for an experiment



- These networks differ *only* in their treewidth!
- Can we design experiments that employ, e.g., scenarios in which the knowledge is structured according to these networks?
- If so, since treewidth is the only variable that is manipulated, indeed treewidth is a source of complexity in the model

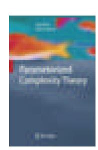
Conclusion

- Despite intractability in general, Bayesian abduction is still a very useful framework for computational cognitive models, but we need to constrain the input to make it tractable (and approximation will not do)
- Theoretical computer science gives us Parameterized Complexity Theory as an off-the-shelf tool to study sources of complexity and constrain the input
- This gives us not only mathematically *sound* models (in addition to psychologically plausible models), but also *empirically testable* hypotheses

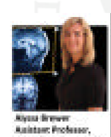
Take home message (reprise)

are to

as



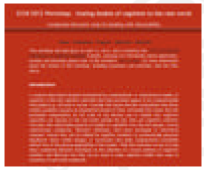
is to



Advertisement

Want to learn more?

I co-organize the workshop
"Scaling models of cognition to the real world"



@ ICCM 2012 conference, April 12th, Berlin

References

1. Mark Blokpoel, Johan Kwisthout, Theo van der Weide and Iris van Rooij (2010). *How Action Understanding can be Rational, Bayesian* and *Tractable*. CogSci 2010.
2. Johan Kwisthout (2011). *Most Probable Explanations in Bayesian Networks: Complexity and Tractability.* International Journal of Approximate Reasoning, 52 (9), 1452 - 1469.
3. Mark Blokpoel, Johan Kwisthout, Todd Wareham, Pim Haselager, Ivan Toni, and Iris van Rooij (2011). *The computational costs of recipient design and intention recognition in communication.* CogSci 2011.
4. Johan Kwisthout, Todd Wareham, and Iris van Rooij (2011). *Bayesian Intractability is not an Ailment that Approximation can Cure*. Cognitive Science, 35 (5), 779 - 784.
5. Johan Kwisthout and Iris van Rooij (2012). Bridging the Gap between Theory and Practice of Approximate Bayesian Inference. *Accepted to the 11th International Conference on Cognitive Modeling*.