Good Alfternoon colloquium, Monday June 20th 2011

Bayesian abduction in Cognitive Science

Johan Kwisthout
Institute for Computing
and Information Sciences

Motivation

- Many computational models of cognitive processes are based on *Bayesian abduction* as its underlying framework
- 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?

✓ Models correlate with empirical observations and intuition...

✗ ...but they are computationally intractable...!

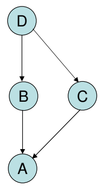
Outline

- Bayesian abduction in a nut shell
  - How do we formalize it as a computational problem
  - Why is it intractable and why would we matter?
- "Now what?" – three ways of dealing with intractability
  - The *doomsday* approach
  - The *hand-waving* approach
  - The *analytical* approach
- Complexity analysis as a research tool
  - Recent research developments
  - New directions and opportunities

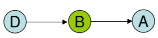
Bayesian abduction

- **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,B,C,D**: a set of stochastic variables
- **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

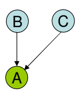


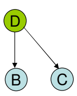
Conditional dependencies



- 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





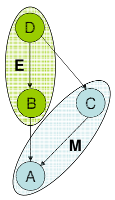
- If **D** is observed, **B** and **C** become independent from each other as **D** is a common cause of **B** and **C**

Calculations in Bayesian Networks

- Take *n* variables with each *m* values. Without independencies, calculations are exponential
- E.g. Pr(**A** = **a**) = $\Sigma_{B,C,D}$ Pr(**A** = **a**, **B**, **C**, **D**)
- We use indepencies to calculate Pr(**A** = **a**) more efficiently
- E.g. Pr(**A** = **a**) = $\Sigma_{B,C}$ Pr(**A** = **a** | **B**, **C**) × Σ_{D} Pr(**C** | **D**) × Pr(**B** | **D**) × Pr(**D**)
- However, in general, this can still be exponential in the number of variables
- Thus calculating the probability distribution of a variable in Bayesian networks is intractable in general

Bayesian abduction problem

- *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 $argmax_{\mathbf{m}}$ Pr(**M** = **m**, **E** = **e**)



- Bayesian abduction or *Most Probable Explanation* happens to be *NP*-hard [1]
- We define a decision variant using an additional rational number *q* and ask whether there exists an **m** such that Pr(**M** = **m**, **E** = **e**) > *q*

Hardness proof

- We show that every *Satisfiability* instance can be reduced to an instance of *Most Probable Explanation* (MPE), in polynomial time, using only a polynomial 'blowup'
- *Satisfiability* (SAT): given a logical formula Φ, is there a truth instantiation to the variables making Φ true?
- This problem is known to be *NP*-complete
- Hence, if there would be a poly time algorithm to MPE, we would also have a poly time algorithm for SAT: hence, *P* would be equal to *NP* which is **HIGHLY** unlikely

NP-Hardness proof of MPE

- Transform a SAT instance to MPE

$$\Phi = \neg(X_1 \vee X_2) \vee \neg X_3$$

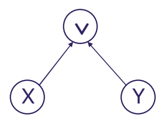
- Is there an instantiation to the variables X_1, X_2, X_3, that satisfies Φ? (Actually there is, say X_1 = T, X_2 = X_3 = F)
- We construct a Bayesian network B_Φ from an instance Φ and designate evidence **e** and a set of variables **M** such that there is an instantiation to **M** (with evidence **e**) with probability > 0 *if and only if* Φ is satisfiable

Hardness proof constructs

- Variables in Φ are nodes with values T, F (uniform probability)

Pr(X = T) = 0.5
Pr(X = F) = 0.5

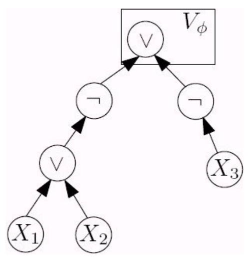
- Operators in Φ are nodes with values T, F (probability table = truth value of logical component)



$Pr(\vee = T | X = T \text{ and } Y = T) = 1$
$Pr(\vee = T | X = T \text{ and } Y = F) = 1$
$Pr(\vee = T | X = F \text{ and } Y = T) = 1$
$Pr(\vee = T | X = F \text{ and } Y = F) = 0$

Hardness proof constructs

$$\Phi = \neg(X_1 \vee X_2) \vee \neg X_3$$



$Pr(V_\Phi = T \wedge X_1 \wedge X_2 \wedge X_3) = 0$
$Pr(V_\Phi = T \wedge X_1 \wedge X_2 \wedge \neg X_3) = 1$
:
$Pr(V_\Phi = T \wedge \neg X_1 \wedge \neg X_2 \wedge \neg X_3) = 1$

Define **E=e** as V_Φ =T and **M** = {X_1, X_2, X_3}. There is an joint value assignment **m** to **M** such that Pr(V_Φ = T ∧ **M** = **m**) > 0 if and only if there is a truth assignment that satisfies Φ!

MPE hardness proof

- Given *evidence* that V_Φ is true, is there a joint value assignment to the variables in X with probability Pr(V_Φ =T, X) > q?
- Any assignment X that satisfies Φ has probability Pr(V_Φ =T, X) = 1
- Any assignment X that does not satisfy Φ has probability Pr(V_Φ =T, X) = 0
- So with *q* ∈ [0, 1>, the answer is yes if and only if Φ is satisfiable! *(note that this holds even if q = 0)*

So what?

- Bayesian abduction is *NP*-hard; so what?
- This means that in *general* there cannot exist a polynomial-time algorithm for solving *arbitrary* instances of Bayesian abduction
- This means that there are 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!"

Outline

- Bayesian abduction in a nut shell
  - How do we formalize it as a computational problem
  - Why is it intractable and why would we matter?
- **"Now what?" – three ways of dealing with intractability**
  - The *doomsday* approach
  - The *hand-waving* approach
  - The *analytical* approach
- Complexity analysis as a research tool
  - Recent research developments
  - New directions and opportunities

The doomsday approach

Bayesian abduction is *NP*-hard

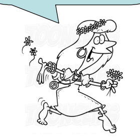
Bayesian models are no good models of the brain

- 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!

Satisficing / approximation?

- We have just seen that determining whether there exists a joint value assignment **m** to **M** with evidence **e** such that Pr(**M** = **m**, **E** = **e**) > *q* is NP-hard **even when q = 0**
- So, not only finding the most probable explanation is *NP*-hard, but *so is finding an explanation with **any** probability higher than zero* [1]
- So the argument "the mind doesn't exactly compute the best explanation, but a sufficiently good enough explanation" is ***still*** not sufficient to solve the intractability problem! [2]

Structural approximation

- OK, finding an approximation with an "*almost as high"* value as the most probable one is *NP*-hard. But what about finding an approximation that ***looks a lot like*** the most probable one?
- Structure approximation of MPE:
  Input: A Bayesian network partitioned in two sets **M** and **E**, and an observation **e** for the variables in **E**, integer *k*
  Output: A value assignment **m'** to **M** with **E** = **e** that differs in at most *k* variables from the most probable explanation **m**
- Bad news again: structure approximation is *NP*-hard [*]

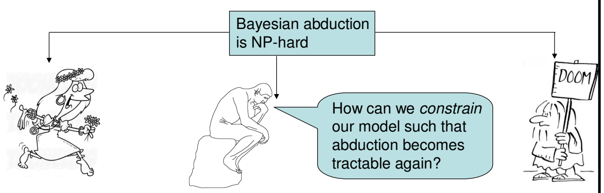
Sketch of a proof – structural approximation

- Say we have an algorithm ***A*** and an integer *k* that computes, for every instance X of MPE, a value assignment that differs in at most *k* variables from the most probable one – we show that this algorithm cannot run in poly time
- We create an instance X^{k+1} consisting of *k* + 1 copies of X. this instance has optimal solution $\mathbf{m}_{\mathbf{opt}}$
- We run ***A*** on X^{k+1} and obtain an approximate solution $\mathbf{m}_{\mathbf{app}}$

Sketch of a proof – continued

- By definition, this solution differs in at most *k* variables from $\mathbf{m}_{\mathbf{opt}}$. As there are k+1 copies of X in X^{k+1}, *at least one of them* has been assigned the optimal solution to the original problem X!
- We can compute the probabilities of each copy in poly time, as there are no free variables, and check which one has the highest probability; we output this value assignment
- Now we have turned our approximation algorithm in an exact solution-algorithm in poly time – so if the exact solver is NP-hard, so must our approximation or else P = NP!

The analytical approach



- 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

- Even when a problem Π is NP-hard in general, it may be the case that there exists 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, the algorithm runs in time $O(n^c)$ which is polynomial

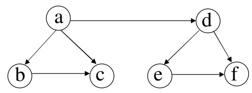
Parameterized complexity analysis of MPE

- 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, …

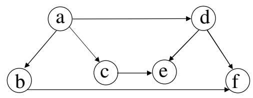
Treewidth

- 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

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!"

Outline

- Bayesian abduction in a nut shell
  - How do we formalize it as a computational problem
  - Why is it intractable and why would we matter?
- "Now what?" – three ways of dealing with intractability
  - The *doomsday* approach
  - The *hand-waving* approach
  - The *analytical* approach
- **Complexity analysis as a research tool**
  - Recent research developments
  - New directions and opportunities

Isotropy and abduction

- **Jerry Fodor**: We are faced with an enormous challenge in Cognitive Science (and AI):

  To make correct abductions, potentially all our knowledge need to be accessible, as we cannot decide beforehand what is relevant and what not. This leads inevitably to intractability, yet we are able to do abductions well enough in practice. Cognitive Science still fails to explain this!

- To remain isotropy (everything may influence everything) it is necessary that our knowledge is stored in a non-modular way, yet this yields computational intractability in general: abduction (Bayesian or otherwise) is *NP*-hard in general

Parameterized complexity to the rescue!

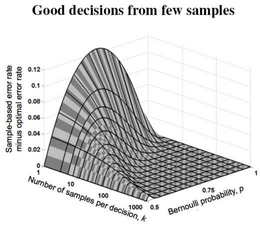
- Abduction is *NP*-hard in general – but it need not be when certain parameters are bounded!
- Joint work with Pim & Iris [4] – what if we assume that knowledge is structured such as to bound treewidth?
- We argue that we can have *both* isotropy *and* tractability if we assume bounded treewidth (and bounded cardinality)
- Parameterized complexity may well be the *truly good idea* that Cognitive Science needs to remedy Fodor's challenge

Alternative notions of Abduction

- Bayesian abduction is typically modeled as an MPE or MAP problem
  - MPE: two-partition in evidence and hypotheses
  - MAP: three-partition, also include intermediate variables that need to be marginalized over
- Set of intermediate variables is HUGE in practice: most of our knowledge is not applicable/relevant for most of our everyday problems
- MPE/MAP may not be the best model of abduction to explain what we observe in practice

Conflicting evidence from CogPsy

- Vul et al. (2009): humans typically make only few samples to decide on the highest utility (e.g. take tunnel / bridge)
- That is OK if the probability of the distribution is skewed



Read as follows:

- Two possible actions that lead to optimal/suboptimal result
- Optimal error is fully rational: I choose the action with the most probable optimal result and I loose 1 – p percent of cases
- Sample based: pick *k* samples and choose the action that led to optimal result in the majority of cases

Explanatory issue with MAP

- There are typically few variables that *normally* play a role in deciding the best explanation for what we observe
  - E.g., to infer why someone is not at his work on the usual time it is relevant to know whether he commutes by train, whether there are traffic jams and whether the trains are late
  - It is *not* relevant to know whether he wore a green or a blue coat on his way to the station
- There are literally millions of such normally irrelevant variables
  - While we *are* able to reason with them, that is not our usual behavior: that is why detective books and series are popular
- Even if it *would* be tractable to do marginalization of all of them, it counters our conceptual notion of best explanation

Thought Experiment: Mr. Jones is late at work

Train delayed?

Got incidentally shot while walking past a bank at 11th Street at 8.03 AM while there was a bank robbery at that very moment?

Thought Experiment

- While trying to explain why Mr. Jones is not at his desk at 8.30 AM, we take some variables into account
  - Does Mr. Jones need to change trains?
  - Is the weather very bad? Is there a strike going on?
- Some variables are typically *not* taken into account
  - Did Mr. Jones walk on the left or right pavement on 11th street?
  - Did he wear a blue or grey coat?
- Only in the awkward coincidence that Mr. Jones happened to be in the wrong place at the wrong moment, they become relevant to explain why he is not at his desk!

Most Parsimonious Explanations

- **Most Parsimonious Explanation** [3]: find an explanation that assumes *as little as possible* of the values of variables that are neither observed nor in the explanation set
- Formalized as finding the explanation that is most probable for the *majority* of value assignment of the intermediate variables – and thus that will likely be picked when only a few samples are taken
- Complexity worst-case comparable with MAP, but with better parameterized approximation algorithms and with possible a better 'explanation' of our cognitive processes

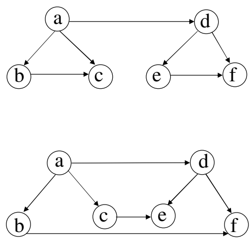
Socrates revisited

- ***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***: "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 cognitive models, but we need to constrain the input to make it tractable
- This gives us not only mathematically *sound* models, but also *empirically testable* hypotheses
- Ongoing and future research:
  - Verify cognitive models, i.e., communicative intention
  - Provide formal results for approximation strategies, e.g., structure approximation or Most Parsimonious Explanation
  - Apply ideas to other areas, e.g., philosophy or neuro-cognition
  - Maybe empirical research with manipulated scenarios

References

1. Johan Kwisthout. *Most Probable Explanations in Bayesian Networks: Complexity and Tractability*. International Journal of Approximate Reasoning, forthcoming.
2. Johan Kwisthout, Todd Wareham, and Iris van Rooij (2011). *Bayesian Intractability is not an Ailment that Approximation can Cure*. Cognitive Science, 35.
3. Johan Kwisthout. Two New Notions of Abduction in Bayesian Networks. *Proceedings of the Benelux Conference on AI, BNAIC 2010, October 25-26, Luxembourg.*
4. Johan Kwisthout, Pim Haselager, and Iris van Rooij. *Abduction, Isotropy and Tractability: Key Ideas for Meeting Fodor's Challenge.* In preparation

* Unpublished research with Iris van Rooij and Todd Wareham