Decision-Theoretic Troubleshooting: Hardness of Approximation

Václav Lín
Institute of Information Theory and Automation, AS CR
Faculty of Management, University of Economics, Prague

Abstract

Decision-Theoretic Troubleshooting is one of the application areas of Bayesian networks. Given a probabilistic model of a malfunctioning device, the task is to find the repair strategy with minimal expected cost. While solvable in polynomial time in simple cases, it is *NP*-hard in general. We show some cases where even computing approximate troubleshooting strategies is *NP*-hard.

Troubleshooting with Bayesian Networks

Heckerman and others at Microsoft research (mid 1990's) [1]:

- printer troubleshooting.

The SACSO project (late 1990's to early 2000's) [2]:

- Collaboration of HP and Machine Intelligence Group at Aalborg University, Denmark.

Recent dissertations – Håkan Warnquist (2011), Thorsten J. Ottosen (2012).

Troubleshooting as a Combinatorial Problem (TS)

The problem is given by

- a set $\mathcal{F} = \{f_1, \dots, f_m\}$ of possible faults,
- a set $\mathcal{A} = \{A_1, \dots, A_n\}$, of available repair actions,
- a probabilistic model $P(\mathcal{F} \cup \mathcal{A})$.

Each fault is fixed by at least one action. Each action A_i

- bears a constant cost $c(A_i)$,
- fixes a subset of faults $F(A_i) \subseteq \mathcal{F}$ with certainty and no other faults,
- can either fail ($A_i =$ no) or succeed in fixing the device ($A_i =$ yes).

The device is repaired by performing the actions in sequence until it is fixed or all actions have been exhausted.

The objective is to find a linear ordering π of actions minimizing the expected cost of repair:

$$ECR(A_{\pi(1)}, A_{\pi(2)}, \cdots) = \sum_{i=1}^{n} p_{\pi(i)} \cdot \underbrace{\sum_{j \le i} c(A_{\pi(i)})}_{\text{cost of first } i \text{ actions}}$$

where

$$p_{\pi(i)} = \underbrace{P\Big(\bigwedge_{j<i} \{A_{\pi(j)} = \text{no}\} \wedge \{A_{\pi(i)} = \text{yes}\}\Big)}_{\text{probability of fixing the device by the } i^{\text{th}} \text{ action}} .$$

ρ-Approximation

Algorithm A is a ρ-approximation algorithm ($\rho > 1$) for a minimization problem L if for all instances x of L we have

$$opt(x) \le A(x) \le \rho \cdot opt(x) ,$$

where

- $A(x)$ is the value returned by A when applied to instance x,
- $opt(x)$ is the optimum of x.

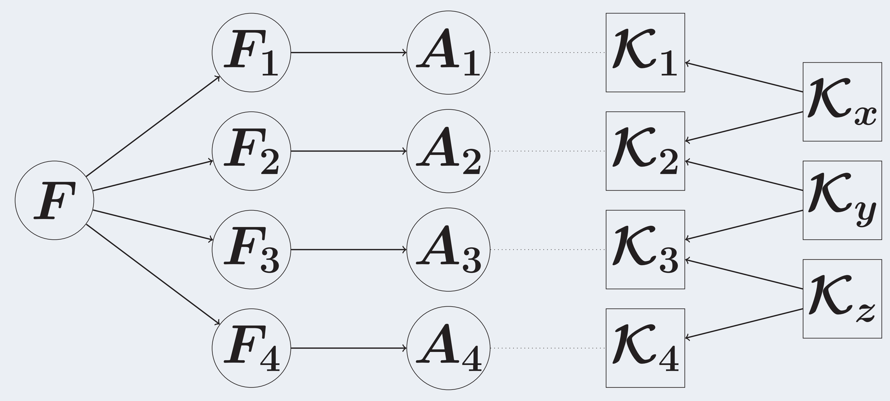
The Min-sum Set Cover Problem (MSSC)

- Input – a finite set U, a collection $\mathcal{C} = \{S \subseteq U\}$.
- Objective – Find a linear ordering π of $\mathcal{C}$ minimizing $\sum_{u \in U} \pi(u)$, where $\pi(u)$ is the index of the first set $S \in \mathcal{C}$ covering element u under the ordering π.

Theorem (Feige, Lovász, Tetali). Unless P=NP, MSSC *has no polynomial-time ρ-approximation algorithm for any $\rho < 4$.*

Variants of the Troubleshooting Problem

- Troubleshooting with dependent actions — the sets of faults $F(A_i)$, $F(A_j)$ addressed by two different actions A_i, A_j may overlap.
- Troubleshooting with dependent faults — several faults may be present in the system at the same time and the faults are not independent.
- Troubleshooting with cost clusters forming an acyclic directed graph — the set of actions is partitioned into disjoint subsets, called *cost clusters*.
  - Cost clusters model the situation when several actions require common initialization work.
  - To access actions within a cluster $\mathcal{K}_i$, we have to pay additional cost, $c(\mathcal{K}_i)$.
  - Once $c(\mathcal{K}_i)$ is paid, the cluser $\mathcal{K}_i$ is *open*, and we can use actions from $\mathcal{K}_i$ at any time.
  - Further, some cost clusters may be accessed only after some other clusters have been opened, as indicated in the figure below. At the right side is the graph of cost clusters. To access action A_2, we have to open clusters $\mathcal{K}_y$ and $\mathcal{K}_2$ (or $\mathcal{K}_x$ and $\mathcal{K}_2$). At the left side is Bayesian network with faults $F_1, \dots, F_4$.



Reductions

Technique: To show that a polynomial-time ρ'-approximation for problem L is *NP*-hard, construct a polynomial-time reduction $\mathcal{R}$ from *MSSC* to L so that a (hypothetical) ρ'-approximation algorithm for L combined with the reduction $\mathcal{R}$ would constitute a ρ-approximation for *MSSC* with $\rho < 4$.

Min-sum Set Cover $\to$ TS with Dependent Actions

$u \in U \longmapsto$ Fault f_u. Distribution $P(\{f_u\}_{u \in U})$ is uniform.
Take the single fault assumption – at any moment of time, at most one fault is present in the system.

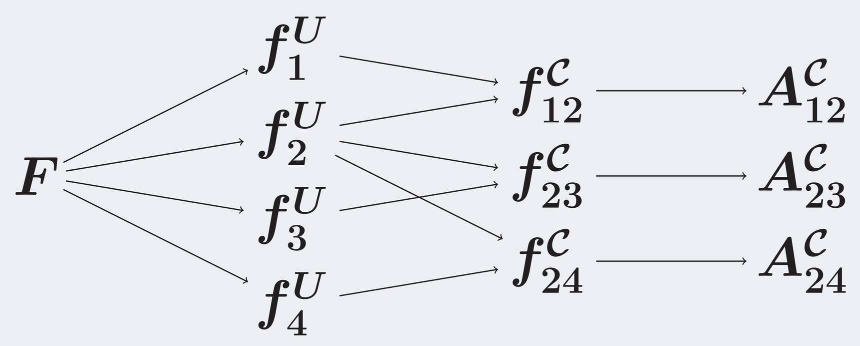
$S \in \mathcal{C} \longmapsto$ Action A_S with cost $c(A_S) = 1$.
Action A_S fixes the set of faults $\{f_u : u \in S\}$

The proof is completed by showing that for an *arbitrary* ordering π of actions (sets) we have

$$ECR(A_{\pi(1)}, A_{\pi(2)}, \cdots) = \frac{\sum_{u \in U} \pi(u)}{|U|}.$$

Min-sum Set Cover $\to$ TS with Dependent Faults

Reduce an instance of *MSSC* to an instance of *TS with Dependent Faults* as indicated in the figure for $U = \{1,2,3,4\}$ and $\mathcal{C} = \{\{1,2\},\{2,3\},\{2,4\}\}$. The f's are faults and A's are actions.
The reduction maps $U \to \{f_u^U : u \in U\}$ and $\mathcal{C} \to \{f_S^{\mathcal{C}}, A_S^{\mathcal{C}} : S \in \mathcal{C}\}$.
Each action has cost one and $P(\{f_u^U : u \in U\})$ is uniformly distributed.



Theorems

In the paper, we prove the following.

Unless P=NP, *the following problems have no polynomial-time ρ-approximation algorithms:*

- *Troubleshooting with dependent actions for any $\rho < 4$,*
- *Troubleshooting with dependent faults for any $\rho < 4$,*
- *Troubleshooting with cost clusters forming a DAG for any $\rho < 3$. The result holds even when the cost cluster graph is bipartite.*

References

[1] Breese, Heckerman: Decision-Theoretic Troubleshooting: A Framework for Repair and Experiment, In: Proceedings of UAI 1996, pp. 124–132.

[2] Jensen, Kjærulff, Kristiansen, Langseth, Skaanning, Vomlel, Vomlelová: The SACSO Methodology for Troubleshooting Complex Systems. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 15, 321–333 (2001).

[3] Ottosen, Jensen: The Cost of Troubleshooting Cost Clusters with Inside Information. In: Proceedings of UAI 2010.

lin@utia.cas.cz