AAAI-24
FEBRUARY 20-27, 2024
Vancouver, Canada

Special Track on Safe, Robust and Responsible AI

A Framework for Data-Driven Explainability in Mathematical Optimization

Kevin-Martin Aigner, Marc Goerigk, *Michael Hartsch*, Frauke Liers, Arthur Miehlich
michael.hartisch@uni-siegen.de

Universität Siegen
University of Siegen
www.uni-siegen.de

Do we need Explainability in Mathematical Optimization? Yes!

- **model transparency:** providing the model alone is insufficient
- **user clarity:** users, especially non-experts, question obtained solutions
- **stakeholder engagement:** planners and workers want transparency
- **operational transparency:** e.g. flight delays, require justification
- **empowering individuals:** workers may question unfavorable shifts
- **consumer rights:** solar panel owners demand power feed explanations

The Framework (informal)

given:
- optimization domain and current instance of interest
- feature functions for solutions and instances
- similarity measure of instances and solutions w.r.t. features
- most similar instances of prior observations

goal:
- find a solution that is
  - close to optimality
  - explainable, i.e.:
    - similar to favorable solutions of similar instances
    - differs from unfavorable solutions of similar instances

Given an optimization instance: Find a high quality solution, that is explainable using most similar instances from the past!

Example: Explain Your Child What to Pack for Summer Camp

- your knowledge of combinatorial optimization will not be convincing
- instead use prior observations:

  "Remember last summer camp when you did not pack a rain coat? You had to stay indoors, while the others were playing outside."
- base arguments on similar events; use positive/negative experiences.
- possible explanation:

  This is a good way to pack your bag since you basically packed the same things last year and were happy with it.

Formalizing the Novel Framework

- $\mathcal{X} \subseteq \mathbb{R}^n$ general optimization domain
- set of instances $\mathcal{I}$, instance $I \in \mathcal{I}$, respective solution space $\mathcal{X}(I)$
- **features spaces** $F_{\mathcal{I}}$ and $F_{\mathcal{X}}$ of instances and solutions
- **features functions** $\phi_{\mathcal{I}} : \mathcal{I} \to F_{\mathcal{I}}$ and $\phi_{\mathcal{X}} : \mathcal{I} \times \mathcal{X} \to F_{\mathcal{X}}$
- metrics $d_{\mathcal{I}} : F_{\mathcal{I}} \times F_{\mathcal{I}} \to \mathbb{R}_+$ and $d_{\mathcal{X}} : F_{\mathcal{X}} \times F_{\mathcal{X}} \to \mathbb{R}_+$ as similarity measures for instances and solutions

Features in Example

- instance features
  - knapsack/bag capacity, profits and weights of items
  - metadata such as weather forecast, season, vacation type
- solution features
  - number of packed items of item group (number of toys, pants, . . .)
  - overall number or weight of packed items

The Framework (formal)

- nominal optimization problem $\min_{\boldsymbol{x} \in \mathcal{X}(I)} f^I(\boldsymbol{x})$
- data on N previous decisions: $(I^i, \boldsymbol{x}^i, \lambda^i)$
  - full description of the instance $I^i \in \mathcal{I}$, employed solution $\boldsymbol{x}^i \in \mathcal{X}(I^i)$, confidence score $\lambda^i \in [-1, 1]$
- most similar instances $S_\epsilon(I) = \{i \mid d_{\mathcal{I}}(\phi_{\mathcal{I}}(I), \phi_{\mathcal{I}}(I^i)) \le \epsilon\}$
- bicriteria optimization model

$$\min_{\boldsymbol{x} \in \mathcal{X}(I)} \left\{ f^I(\boldsymbol{x}),\ \sum_{i \in S_\epsilon(I)} \frac{\lambda_i d_{\mathcal{X}}\left(\phi_{\mathcal{X}}(I, \boldsymbol{x}), \phi_{\mathcal{X}}(I^i, \boldsymbol{x}^i)\right)}{1 + \beta d_{\mathcal{I}}\left(\phi_{\mathcal{I}}(I), \phi_{\mathcal{I}}(I^i)\right)} \right\}$$

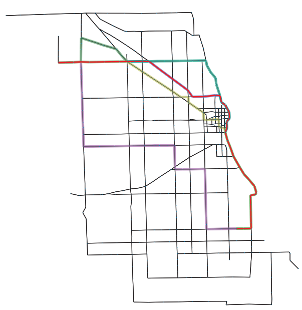
- weighted sum formulation

$$\min_{\boldsymbol{x} \in \mathcal{X}(I)} \alpha f^I(\boldsymbol{x}) + (1-\alpha) \sum_{i \in [N]} \tilde{\lambda}_i d_{\mathcal{X}}(\phi_{\mathcal{X}}(I, \boldsymbol{x}), \phi_{\mathcal{X}}(I^i, \boldsymbol{x}^i))$$

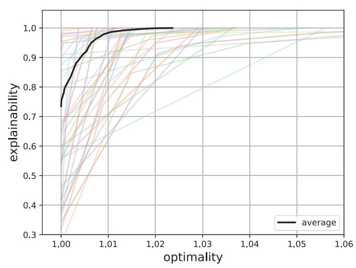
Theoretical Results

- solving the weighted sum formulation is NP-hard and not approximable, already in easy settings
- weighted sum formulation can be solved in polynomial time if nominal problem can be solved in polynomial time for arbitrary costs and $\Phi_{\mathcal{X}}$ being the Hamming distance
- if $\mathcal{X}$ is the set of all s-t-paths in a graph: weighted sum formulation is NP-hard and not approximable.

Experiments: Explainable Shortest Path



- city of Chicago: 538 nodes and 1287 edges
- real-world data of busses
- 4363 observed scenarios with instance features:
  - (average) edge velocities
  - date and time
- solution features: traversed edges
- 50 random s-t pairs, explained by optimal solutions in observed scenarios ($\lambda_i = 1$)
- solve weighted sum formulation for several values of α
- relative optimality score $\left(\frac{\text{optimality value}}{\text{best optimality value}}\right)$
- relative explainability score $\left(\frac{\text{lowest explainability value}}{\text{explainability}}\right)$



Best possible explainable solution has increase in objective value of less than 3% on average

Key Takeaway

The cost of enforcing explainability can be very small!

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg
Universität Siegen
UNIVERSITY OF PASSAU
TRR 154 Mathematische Modellierung, Simulation und Optimierung am Beispiel von Gasnetzwerken