Multi-Stage Predict+Optimize for (Mixed Integer) Linear Programs

Xinyi Hu [1], Jasper C.H. Lee [2], Jimmy H.M. Lee [1], Peter J. Stuckey [3]

1. The Chinese University of Hong Kong
2. University of California, Davis
3. Monash University

Problem Setting

Aim to Solve:
Optimization problems (OPs) with unknown parameters

Given: related information of unknown parameters

- Features
- Historical data (features, true parameters)

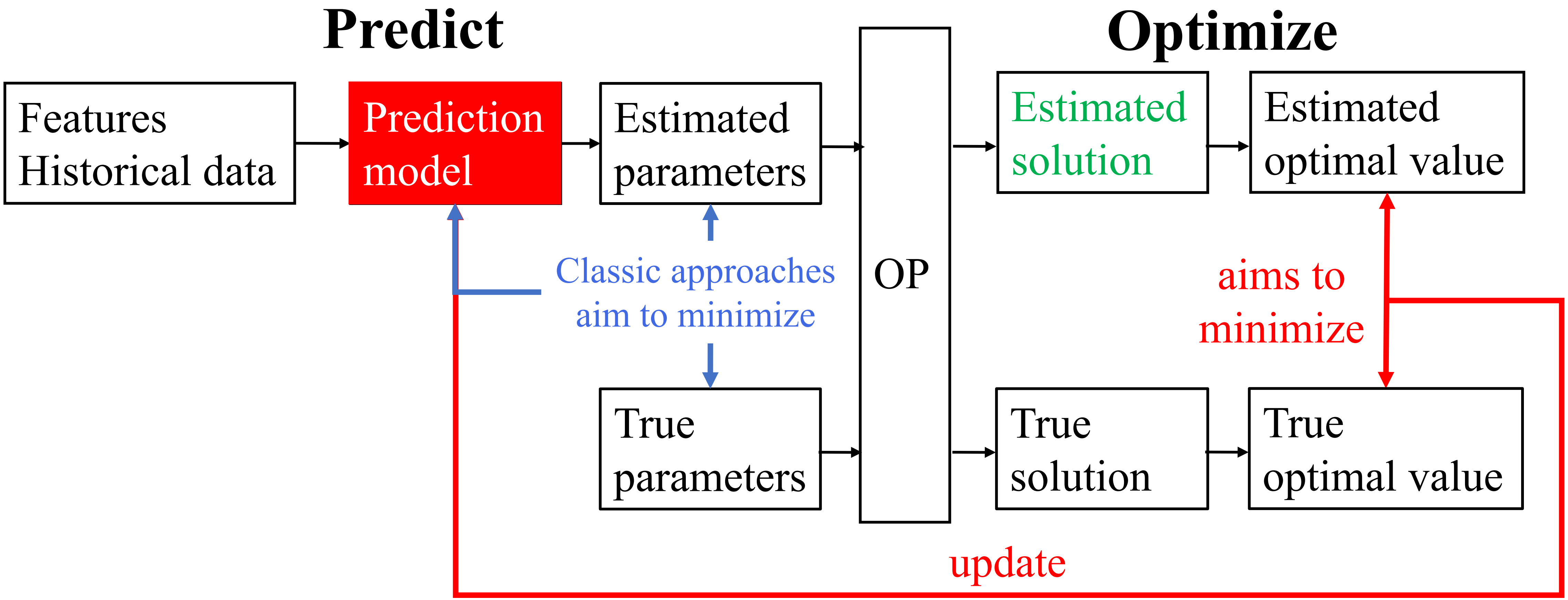
Predict unknown parameters and solve the OP

Predict+Optimize

[Elmachtoub and Grigas, Management Science 2022]

Pipeline: Predict+Optimize

Goal
Good estimated solutions under true parameters



Motivation

Motivation

- Prior frameworks: ***all*** unknown parameters are ***revealed simultaneously***
- Excluding applications
  - Unknown parameters are gradually released
  - New decisions need to be made across many stages

Our goal: OPs with ***gradually revealed*** unknown parameters

Motivating Scenario

OP with *gradually revealed* unknown parameters

Nurse rostering problem:
Every Wednesday, make nurses schedule for the next week (5 days)
Min: total costs for hiring nurses
Subject to: meet patients' demands ? → Unknown

Hospital:
make weekly schedule under
unknown patients' demands
$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) \in \mathbb{R}^5$

Appointment system

Very unfriendly to patients

Reservations closing at some time points → Prior works':
reservations for the whole next week are closed this Sunday

Contributions

Motivation

- Prior frameworks: ***all*** unknown parameters are ***revealed simultaneously***

Contributions

- *Multi-Stage Predict+Optimize Framework*
    - The ***first*** P+O framework for OPs with ***gradually revealed*** unknown parameters
- ***Three*** *End-to-End Training* ***Algorithms***

[Contribution 1] Multi-Stage Predict+Optimize Framework

Predicted demands $\widehat{\boldsymbol{\theta}}^{(0)} = \left(\hat{\theta}_1^{(0)}, \dots, \hat{\theta}_5^{(0)}\right)$

True demands (newly revealed) None

True demands (previously revealed) None

Decisions

Stage 0 solution $\widehat{\boldsymbol{x}}^{(0)}$:
Solve original OP using ***predictions*** $\widehat{\boldsymbol{\theta}}^{(0)}$

Rest
Work

Nurse ID	Mon	Tue	Wed	Thur	Fri
1					
2					
3					

Stages (Time points)

Stage 0 (This Wednesday)

More patient-friendly setting:

- Reservations are closed the day before each working day
- Update future predictions and future schedule everyday

[Contribution 1] Multi-Stage Predict+Optimize Framework

Predicted demands — $\hat{\boldsymbol{\theta}}^{(t)} = \left(\hat{\theta}^{(t)}_{t+1}, \ldots, \hat{\theta}^{(t)}_{5}\right)$: new predictions for the Day $t+1$ and after

True demands (newly revealed) — θ_t: patient demand for the next day, i.e., the Day t

True demands (previously revealed) — $\theta_1, \ldots, \theta_{t-1}$: patient demands for the previous $t-1$ days already reveal

Decisions

Rest
Work

Stages (Time points)

Stage 0 (This Wednesday)

Stage t (Day $t-1$'s night in next week)

[Contribution 1] Multi-Stage Predict+Optimize Framework

Predicted demands

True demands
(newly revealed)

True demands
(previously revealed)

Decisions

Stage t solution $\hat{\boldsymbol{x}}^{(t)}$:
Solve Stage t OP using ***revealed demands*** $\theta_1, \ldots, \theta_t$ and ***new predictions*** $\hat{\boldsymbol{\theta}}^{(t)}$

- Stage t OP modifies the original OP by:
  - Adding constraints that the first t day's schedule cannot be changed
  - Adding a penalty term to the objective

Rest
Work

Nurse ID	Mon	…	Day t	…	Fri
1					
2					
3					

Cannot change since:
- Already happened
- Too late to modify

Stages (Time points)	Stage 0 (This Wednesday)	Stage t (Day $t-1$'s night in next week)

[Contribution 2] End-to-End Training Algorithms

Contributions

- *Multi-Stage Predict+Optimize Framework*
  - The ***first*** P+O framework for OPs with ***gradually revealed*** unknown parameters

- *Three End-to-End Training Algorithms*
  - *Baseline*:
    - A straightforward generalization of the prior work [Hu et al., NeurIPS 2023]
    - Only trains a *single neural network* in Stage 0
  - ***Sequential Coordinate Descent (SCD)***
  - ***Parallel Coordinate Descent (PCD)***

Core idea:
Trains one neural network (NN) and update predictions ***per stage***

Key Experiment Results

Method list:

Proposed

- ***Baseline***
- ***SCD***
- ***PCD***

BAS: ***B***est results obtained ***A***mong all ***S***tandard regression methods

Evaluation:

"***Win rate***" tables: the number of simulations where the proposed beat BAS

Price group	Stage num	Baseline beats BAS	SCD beats BAS	PCD beats BAS
Low-profit	4	93.33%	96.67%	86.67%
	12	73.33%	100.00%	90.00%
High-profit	4	66.67%	96.67%	73.33%
	12	76.67%	100.00%	80.00%

Production and sales problem.

Capital	Stage num	Transaction factor	Baseline beats BAS	SCD beats BAS	PCD beats BAS
25	4	0.01	53.33%	86.67%	73.33%
		0.05	66.67%	90.00%	86.67%
		0.1	70.00%	93.33%	90.00%
	12	0.01	66.67%	93.33%	83.33%
		0.05	80.00%	96.67%	93.33%
		0.1	83.33%	100.00%	96.67%
50	4	0.01	60.00%	80.00%	66.67%
		0.05	66.67%	93.33%	83.33%
		0.1	70.00%	96.67%	90.00%
	12	0.01	70.00%	83.33%	83.33%
		0.05	73.33%	90.00%	86.67%
		0.1	76.67%	100.00%	90.00%

Investment problem

Extra nurse payment	Baseline beats BAS	SCD beats BAS	PCD beats BAS
15	70.00%	70.00%	70.00%
20	73.33%	86.67%	80.00%
25	73.33%	96.67%	83.33%
30	73.33%	86.67%	76.67%

Nurse rostering problem

Takeaways: Solution Quality

- SCD outperforms BAS in almost all simulations
- PCD and Baseline outperform BAS in most simulations

Contributions

- Multi-Stage Predict+Optimize Framework
  - The ***first*** P+O framework for OPs with ***gradually revealed unknown parameters***

- ***Three*** End-to-End ***Training Algorithms***
  - Baseline
  - Sequential Coordinate Descent
  - Parallel Coordinate Descent

xyhu@cse.cuhk.edu.hk