Some Progress Towards Artificial Intelligence for Operations Research

Mahdi Mostajabdaveh
Huawei Vancouver Research Center

The Overall Process of Operations Research Projects

The OR process requires close collaboration among experts from **business, operations research,** and **mathematical programming** fields：

- ✓ **Business Experts**： They provide domain knowledge, including the definition of objectives, decision content, and requirements, as well as the evaluation and usage of solutions.
- ✓ **Operations Research Experts**： They build mathematical optimization models based on the business demands.
- ✓ **Mathematical Programming experts**: They implement efficient algorithms to solve mathematical optimization problems.

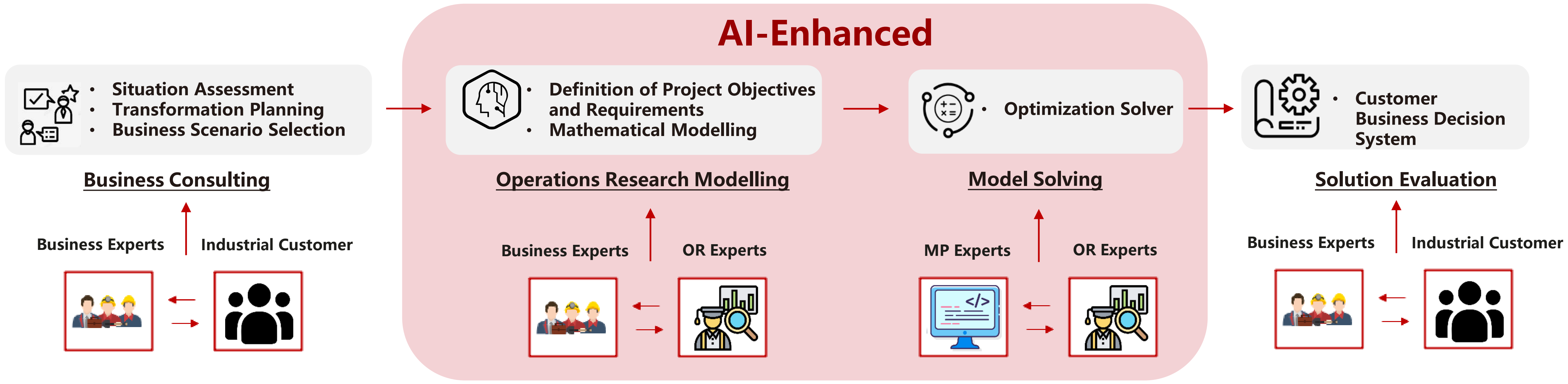


Table of Contents

NL4Opt Competition in NeurIPS2022

Formulating Optimization Problems Based on Their Natural Language Descriptions

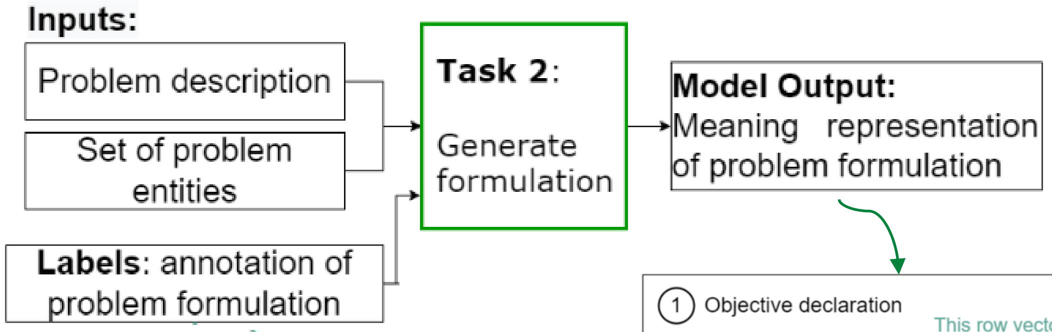
Competition Motivation:

- Enabling users with limited knowledge of OR to create optimization models and solve them.
- Serving as a first step to tackle challenges faced in converting NL to optimization models.
- Addressing the limited learning resources by open-sourcing our dataset and encouraging others to contribute as well.

Two Sub-tasks:

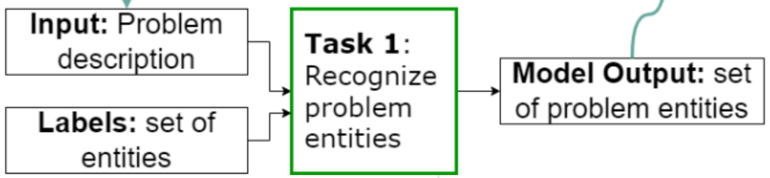
1. Recognition of optimization problem entities

2. Generation of problem formulation.



Your client has $60,000 available to invest for a one-year term. The money can be placed in a trust yielding a 7% return or in a savings account yielding a 2% return. Based on your client's investment goals, you advise her that at least 15% of the investment be placed in the trust. Given her risk profile, she also requests that the money placed in savings should not exceed 60% of her total investment. How much should your client invest in each so as to maximize her return?

[{ "text": "60,000", "label": "limit", "start_char": 17, "end_char": 23},
{ "text": "available", "label": "constraint_direction", "start_char": 24, "end_char": 33},
{ "text": "trust", "label": "variable", "start_char": 94, "end_char": 99},
...
{ "text": "maximize", "label": "objective_direction", "start_char": 400, "end_char": 409},
{ "text": "return", "label": "objective_name", "start_char": 413, "end_char": 433}]



① Order mapping of variable entity mentions

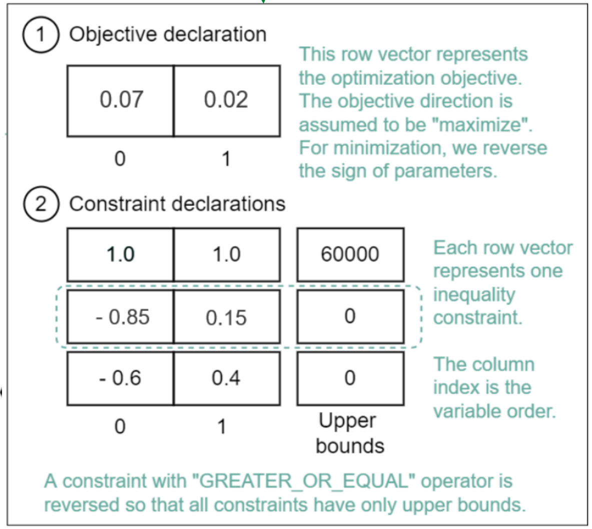
{ "trust" : 0,
"savings account": 1,
"savings": 1 }

② Objective declaration

{ "type": "objective","name": "return",
"direction": "maximize",
"terms": { "trust": 0.07 , "savings": 0.02 }}

③ List of constraint declarations

[{ "type": "sum", "direction": "available",
"limit": "60000", "operator": "LESS_OR_EQUAL"},
{ "type": "ratio", "direction": "at least", "limit": "15%",
"var": "trust", "operator": "GREATER_OR_EQUAL"},
{ "type": "ratio", "direction": "not exceed",
"limit":"60%", "var": "savings",
"operator": "LESS_OR_EQUAL"}]



NL4Opt Competition in NeurIPS2022

Formulating Optimization Problems Based on Their Natural Language Descriptions

❑ Competition Dataset :

Total Samples: 1101 expert annotated Linear Programming Word Problems across 6 domains.

Evaluation Focus: Generalizability to unseen domains.

❑ Some statistics of the competition:

150+ teams registered

300+ submissions evaluated

Total of **28 sets of valid entries**

- 19 for subtask 1
- 9 for subtask 2

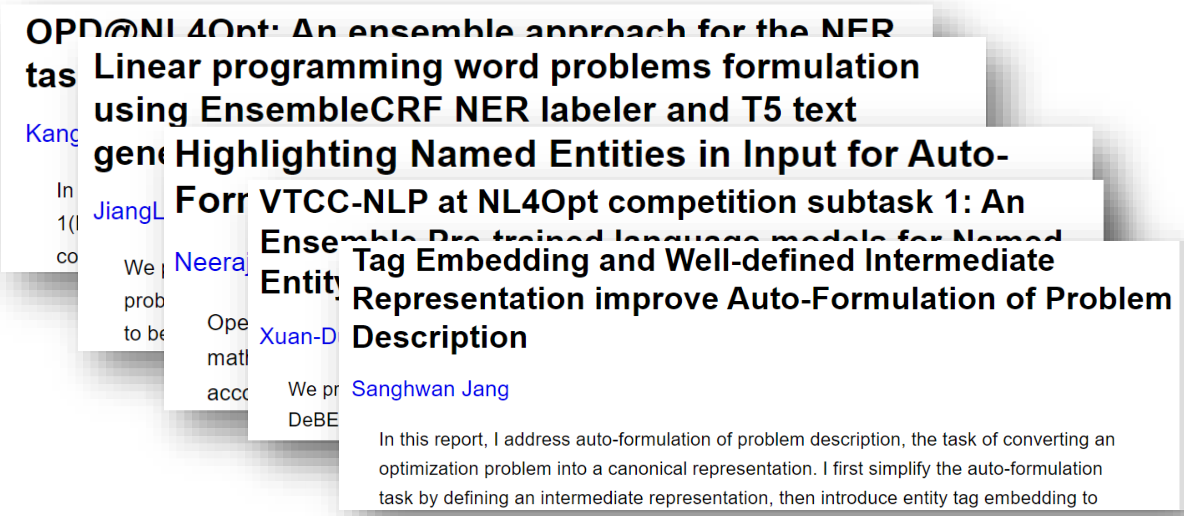
Subtask 1 Winner accuracy was **0.933**, improved **3.3%** over baseline

Subtask 2 Winner accuracy was **0.899,** improved **28.9%** over baseline

❑ **Demographics of submitting teams:**

- Industry (60%), University (25%), Unknown (15%)

Five papers archived by participants



❑ Experiment with LLMs (gpt-3.5 model)

- Without per-training or fine-tuning
- Achieved **0.897** accuracy on test set for the combined task (without receiving intermediate entity tagging)

Ramamonjison, R ... & Zhang, Y. (2023). NL4Opt Competition: Formulating Optimization Problems Based on Their Natural Language Descriptions. *arXiv preprint arXiv:2303.08233*.

New dataset for real-world problems

NL4OPT Dataset

- ❑ **The NL4OPT (level 1) problems are far from the real problems**
- ❑ **Level 2 problem dataset**

Description Abstraction

- No problem name,
- No OR jargon,
- No numerical value,
- No math symbols
- Specific Context

Complexity of the Mathematical Model:

- LP, MILP, QP
- Numbers of Set and Variable <= 5
- Number of Parameters and constraints <= 8
- Covering 15 application domain

Description

Due to the recent surge in COVID-19 cases centered around Yellow Park, as the city's transportation manager, you have been asked to create a plan. The Circle Plaza station, a bustling hub close to multiple tourist attractions, is a region at risk of becoming a super-spreader site due to its high amount of foot traffic. To prevent the spread of the virus from Yellow Park to the Circle Plaza station, the city has decided to temporarily shut down some subway lines. This was proposed to ensure that no subway route exists between Yellow Park and Circle Plaza. While public health is important, shutting down subway lines has its repercussions. There is an impact on city finances. Specifically, impacting ticket sales, staffing, and operational costs. Therefore, you need to make a decision that ensures public safety while also being economically feasible. You have estimated the financial impact associated with closing each track. Your task is to identify which subway routes to shut down, minimizing the overall costs, while ensuring that travellers cannot travel by subway between Yellow Park and the Circle Plaza.

Problem Specifications

- Set of subway stations
- Set of subway routes each connecting two sequential stations
- The cost to shut down a subway route
- Decision to close a specific subway route between a pair of stations
- Decision to include a specific station in the set of stations remain connected to Yellow Park station
- Minimize the cost of closing subway routes to make sure that people cannot travel by subway between Yellow Park and Circle Plaza.
- The Yellow Park station (the Source) is definitely included in the set of stations on the Yellow Park station side, and Circle Plaza station (the Sink) is not.
- If the subway route between two specific stations is closed, then the head station must be included in the set of stations on the Yellow Park side and the tail station is not.

Model Code in Zimpl

```
# Sets:
## Set of subway stations
set STATION;
## Set of subway routes each connecting two sequential stations
set ROUTE[STATION, STATION];
# Parameters:
## The cost to shut down a subway route
param cost[ROUTE];
# Variables:
## Decision to close a specific subway route between a pair of stations
var Z[ROUTE] binary;
## Decision to include a specific station in the set of stations remain connected to
Yellow Park station
var Y[STATION] binary;
# Objective:
## Minimize the cost of closing subway routes to make sure that people cannot travel
by subway between Yellow Park and Circle Plaza.
minimize obj: sum <i, j> in ROUTE do cost[i, j] * Z[i, j];
# Constraints:
## The Yellow Park station (the Source) is definitely included in the set of stations
on the Yellow Park station side, and Circle Plaza station (the Sink) is not.
subto source_sink: Y[first(STATION)] - Y[last(STATION)] >= 1;
## If the subway route between two specific stations is closed, then the head
station  must be included in the set of stations on the Yellow Park side and the tail
station is not.
subto route_cut: forall <(i, j)> in ROUTE do Z[i, j] - Y[i] + Y[j] >= 0;
```

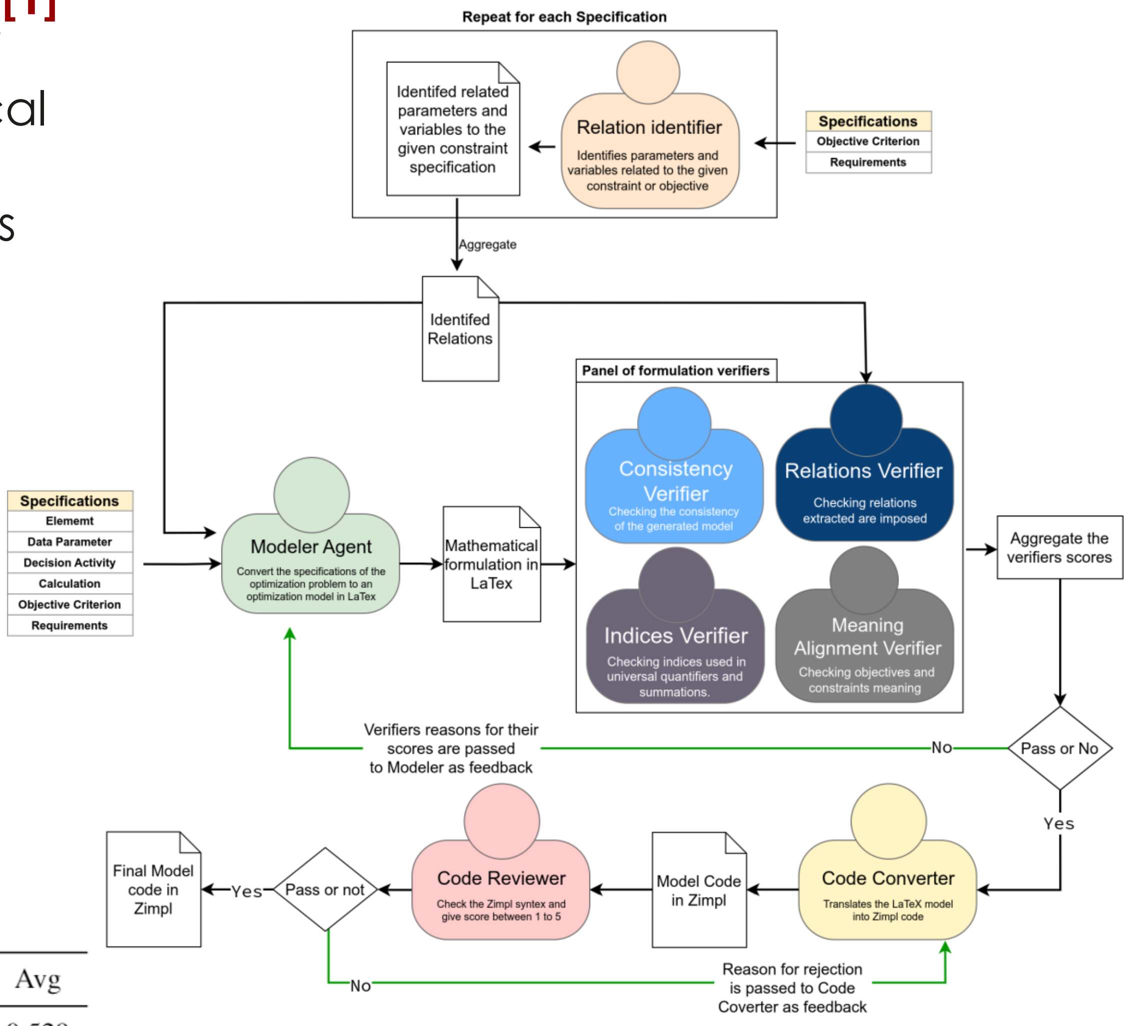
Dataset contains 70 instances design and verified by OR experts

Optimization modeling and verification from problem specifications using a multi-agent multi-stage LLM framework[1]

- LLMs are struggling to accurately generate mathematical models for real-world optimization problems.
- Our multi-agent, multi-stage framework is a step towards enhancing LLM-based methods.

System main characteristics:

- Seven LLM-based agents, each specialized in a specific task.
- Two-stage generation: first from a natural language description to a mathematical model in LaTeX, and then from LaTeX to model code.
- Inter-agent communication to resolve errors.
- Multi-verifier with a voting mechanism to verify the mathematical model.



Experiment on the Level-2 Dataset

Strategy	Multi-turn	Spec input	Desc input	Set	Param	Var	Obj	Constraint	Avg
				Component Exact-match Accuracy					
DESC2MODEL			✓	0.821	0.633	0.448	0.200	0.108	0.529
SPEC2MODEL		✓		**1.000**	0.889	**0.829**	0.586	0.472	0.747
MULTI-TURN	✓	✓		**1.000**	0.832	0.770	0.500	0.426	0.712
MULTI-TURN + DESC	✓		✓	**1.000**	**0.893**	0.789	0.600	0.458	0.751
MULTI-TURN + SPEC	✓	✓		**1.000**	0.881	0.789	0.571	0.463	0.746
OUR APPROACH		✓		**1.000**	0.873	0.786	**0.804**	**0.689**	**0.808**

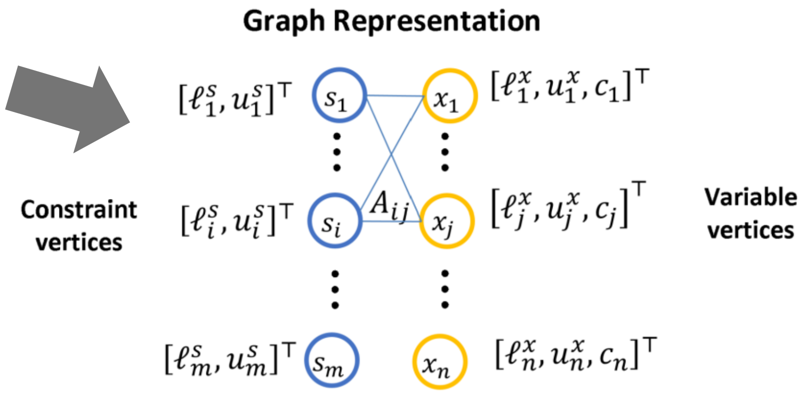
[1] - Mostajabdaveh, M., Yu, T. T., Ramamonjison, R., Carenini, G., Zhou, Z., & Zhang, Y. (2024). Optimization modeling and verification from problem specifications using a multi-agent multi-stage LLM framework. *INFOR: Information Systems and Operational Research*, 1-19.

Human-aligned Evaluation for mathematical models (COLING 2024)

Our method

- Convert the LP/MILP to a Bipartite Graph (which is **permutation invariance**)
- Use Graph Edit Distance (GED) as the evaluation metric.
- GED is defined by the minimum-cost sequence of basic edit operations to transform one graph into another by means of insertion, deletion and substitution of vertices and/or edges.

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.} \quad & \boldsymbol{\ell}^s \leq \mathbf{A}\mathbf{x} \leq \mathbf{u}^s \\ & \boldsymbol{\ell}^x \leq \mathbf{x} \leq \mathbf{u}^x \end{aligned}$$



Canonical Metric

- Based on the declaration-level matching between hypothesis and reference model.
- **Issue:** Not robust to the altered order of variables:

$$a\cdot X + b\cdot Y \leq c$$
$$b\cdot Y + a\cdot X \leq c$$

Solver Executable Metric [3]

- Comparing the optimal solutions between hypothesis and reference models.
- **Issue:** Models with the same optimal solution (or infeasible) are not distinguishable.

Correlation with Human Evaluation

Metrics	C-Match	F-Match
Execution	9 / 289	716 / 1734
Canonical	64 / 289	1336 / 1734
Ours	**178 / 289**	**1641 / 1734**

C-Match measures the percentage of instances where the human and automatic ranking lists exactly match.
F-match decomposes ranking lists into individual ranking pairs and then calculates the match rate at the pair level.

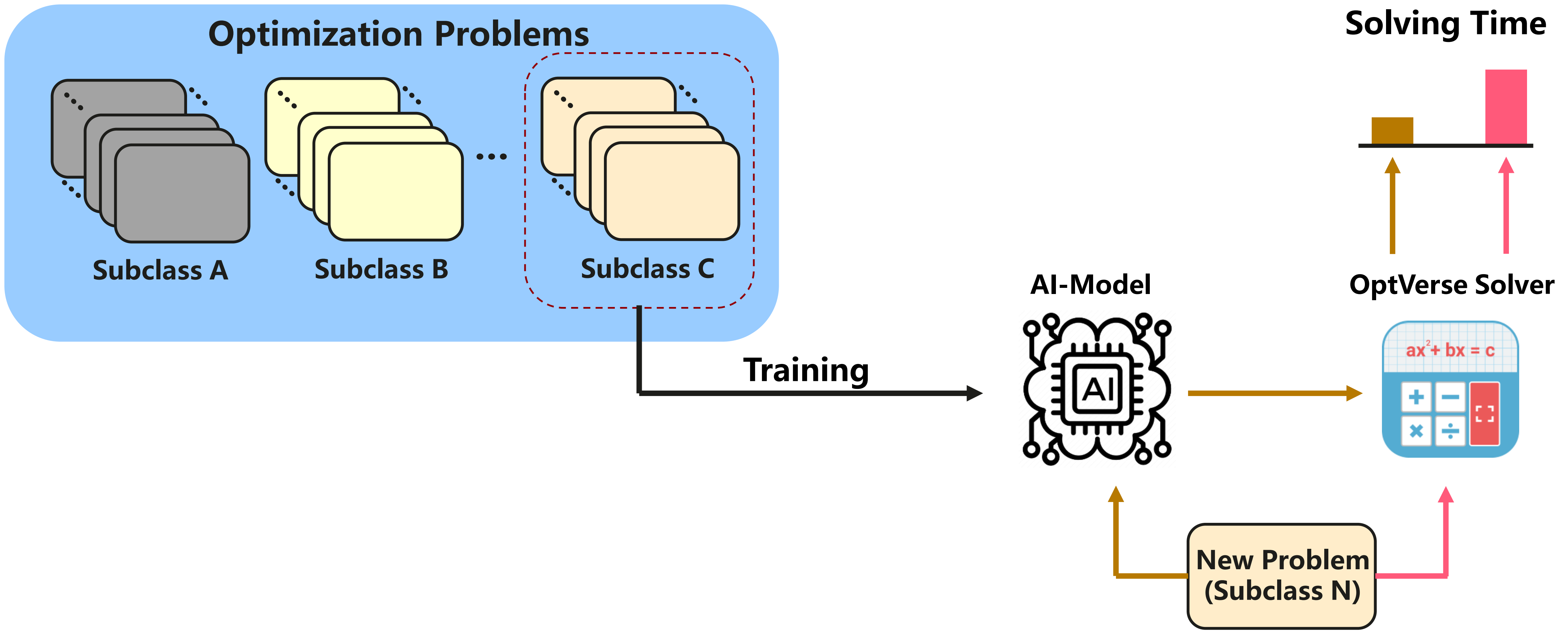
Xing, L., et al. (2024, May). Towards Human-aligned Evaluation for Linear Programming Word Problems. In *Proceedings of LREC-COLING 2024,* pp. 16550-16556.

Table of Contents

AI-Enhanced Optimization Solver

No Free Lunch Theorem [Wolpert and Macready, 1997]

- All optimization algorithms perform equally well when their performance is averaged over all possible objective functions.
- Specialization to a subclass of problems is in fact the only way that improved performance can be achieved in general.



Smart Initial Basis Selection for Linear Program (ICML 2023)

- Linear program (LP) has been a fundamental aspect of various industrial domains, such as airplane scheduling and product planning.
- Simplex method is a pioneering method for solving LP. It starts with an initial basis $\mathcal{B}^{(0)}$ and routinely pivots to a neighboring basis with improvement until reaching an optimal basis $\mathcal{B}^*$. Its **efficiency** is greatly affected by the **initial basis**.
- Existing rule-based basis selection strategies leverage linear algebra heuristics. **We propose a learning-based approach for scenarios where LP problems are correlated**, *e.g.*, the airport handling numerous similar hourly flight scheduling problems every day.

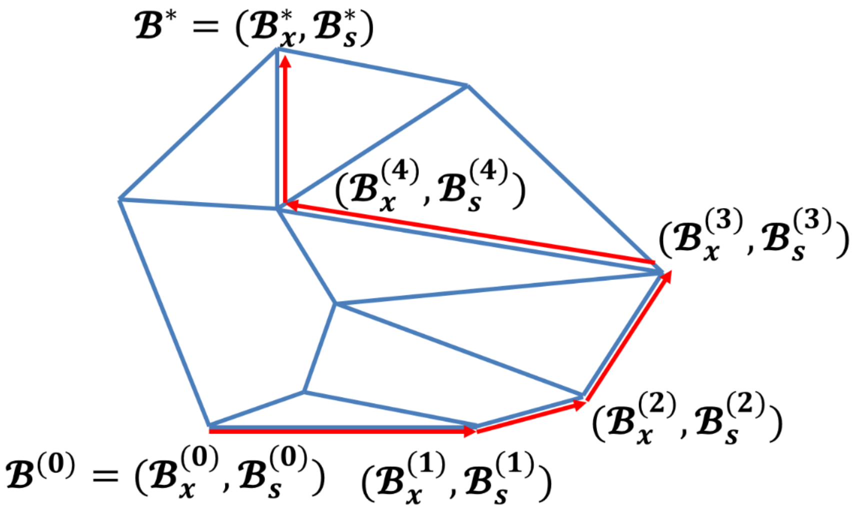


Figure 1 Illustration of Simplex algorithm

LP Problem

$$
\begin{array}{cc}
\min\limits_{x \in \mathbb{R}^n, s \in \mathbb{R}^m} & c^T x \\
\text{s.t.} & Ax = s \\
& \ell^x \leq x \leq u^x \\
& \ell^s \leq s \leq u^s
\end{array}
$$

Smart Initial Basis Selection for Linear Program (ICML 2023)

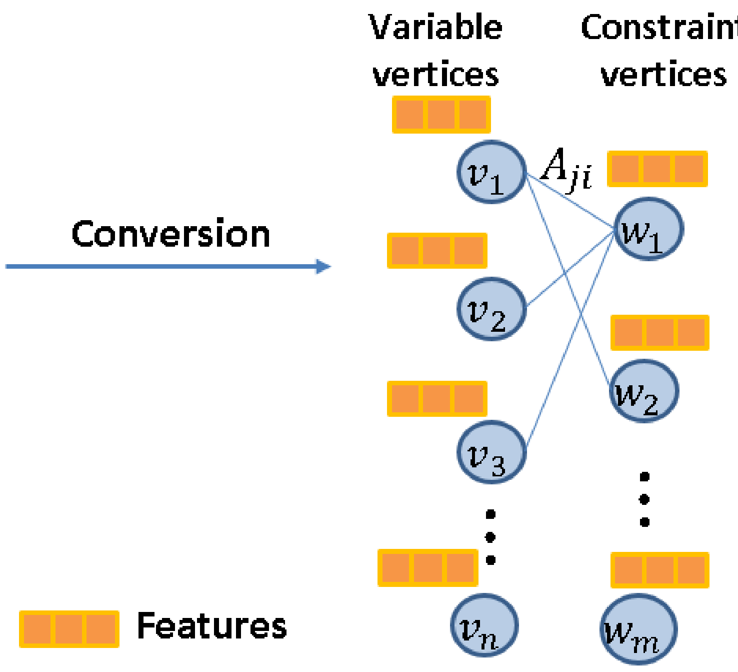
Training Stage

- Represent an LP as a bipartite graph
- Construct labels for solved LPs, build a trainset
- Train a Graph Neural Network (GNN) with knowledge-based masking
  - Knowledge-based masking is integrated into GNN
    - For non-basic entries, make sure the produced probabilities satisfied their feasibility
    - Achieved by adding large penalty to the logits of unreachable bounds
- Use crossentropy to measure the dismatch

Linear Program

$$\begin{aligned} \min_{x\in\mathbb{R}^n, s\in\mathbb{R}^m} \quad & c^T x \\ \text{s.t.} \quad & Ax = s \\ & \ell^x \le x \le u^x \\ & \ell^s \le s \le u^s \end{aligned}$$

Graph Representation



Smart Initial Basis Selection for Linear Program (ICML 2023)

Inference Stage

- GNN inference: predicts basis status probability vectors

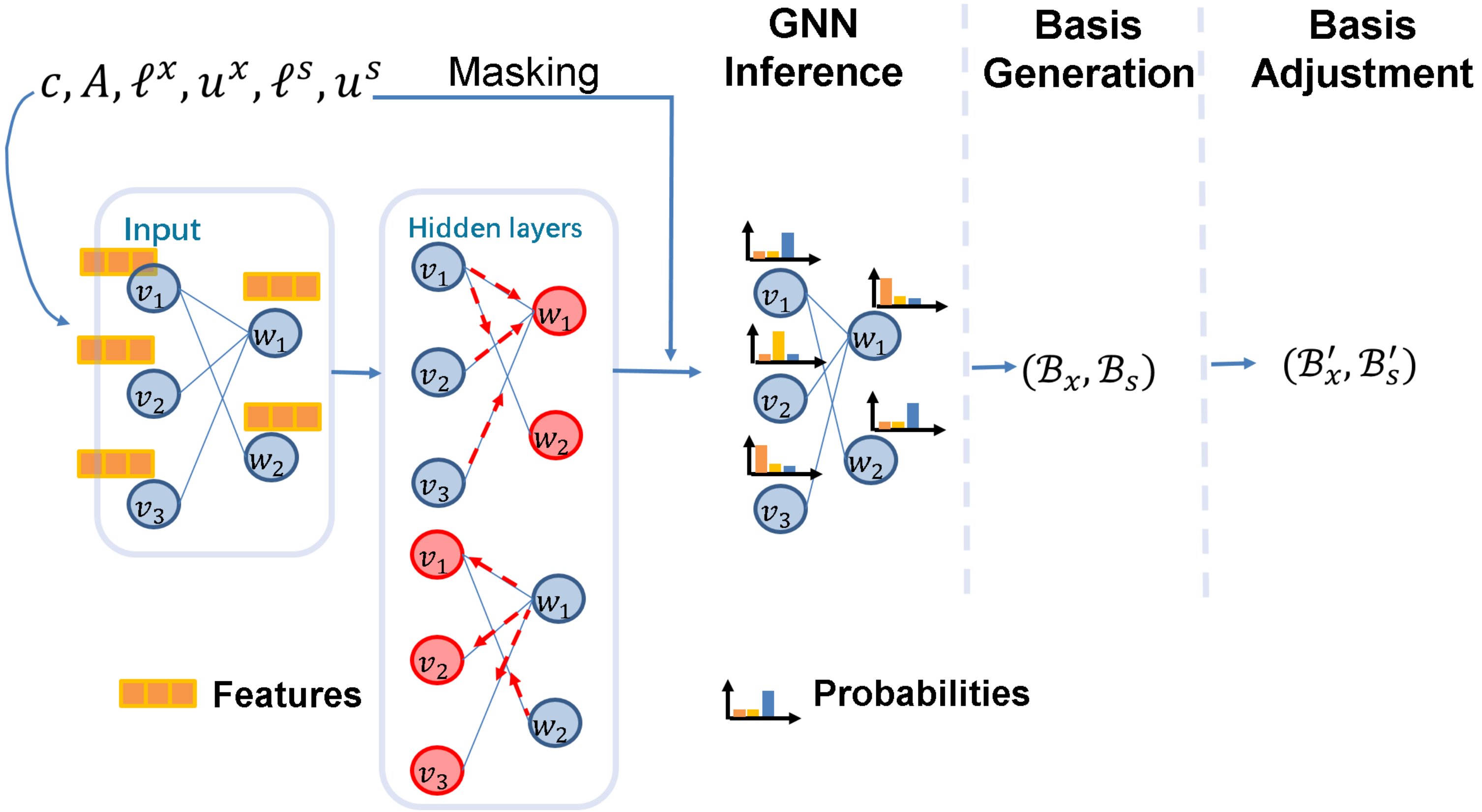
$$\{p_{x,i}, p_{s,j} \mid i \in [n], j \in [m]\}$$

- Basis generation: select top-m constraint and variable indices as basis

$$(\mathcal{B}_x, \mathcal{B}_s)$$

- Basis adjustment: make sure the basis is valid by trying to factorize the corresponding constraint matrix

$$[A_{\mathcal{B}_x} \quad -I^m_{\mathcal{B}_s}]$$



Smart Initial Basis Selection for Linear Program (ICML 2023)

Experiment Results

The proposed method (GNN) is compared to three rule-based basis-selection strategies on six benchmark datasets .

Pros. The traditional basis-selection strategies overlook the information in past solved problems. In contrast, our proposed learning-based strategy consistently outperforms traditional strategies, especially in scenarios with correlated LP problems like daily supply-chain planning.

Cons. We also extensive explored the limitation of our proposed method and shows its efficiency decreases when LP problems are largely uncorrelated.

	Iterations					Time (s)				
Dataset	DEFAULT	CA	CA-MPC	CA-ANG	GNN	DEFAULT	CA	CA-MPC	CA-ANG	GNN
LIBSVM	$14.9K_{\pm 9.5K}$	$14.9K_{\pm 9.5K}$	$21.0K_{\pm 4.8K}$	$15.2K_{\pm 1.1K}$	$\mathbf{9.1K}_{\pm 3.1K}$	$16.6_{\pm 10.0}$	$16.7_{\pm 10.0}$	$27.9_{\pm 12.4}$	$28.3_{\pm 2.2}$	$\mathbf{11.0}_{\pm 3.7}$
MIRP	$40.3K_{\pm 23.3K}$	$34.8K_{\pm 20.2K}$	$36.7K_{\pm 20.8K}$	$39.6K_{\pm 22.7K}$	$\mathbf{25.9K}_{\pm 16.9K}$	$22.1_{\pm 23.3}$	$21.4_{\pm 22.5}$	$18.6_{\pm 16.9}$	$21.6_{\pm 20.9}$	$\mathbf{15.4}_{\pm 15.7}$
STOCH	$75.3K_{\pm 4.3K}$	$52.5K_{\pm 4.8K}$	$48.7K_{\pm 5.2K}$	$53.3K_{\pm 1.7K}$	$\mathbf{31.8K}_{\pm 14.3K}$	$44.6_{\pm 11.8}$	$61.3_{\pm 12.3}$	$51.3_{\pm 12.4}$	$53.2_{\pm 8.5}$	$\mathbf{42.7}_{\pm 30.0}$
GEN	$2.4K_{\pm 225.0}$	$2.4K_{\pm 225.0}$	$2.4K_{\pm 225.0}$	$2.4K_{\pm 225.0}$	$\mathbf{552.8}_{\pm 642.9}$	$1.3_{\pm 0.2}$	$1.4_{\pm 0.2}$	$1.4_{\pm 0.3}$	$1.4_{\pm 0.3}$	$\mathbf{0.5}_{\pm 0.5}$
SC-1	$272.3K_{\pm 151.9K}$	$158.9K_{\pm 89.1K}$	$266.9K_{\pm 148.5K}$	$269.2K_{\pm 151.5K}$	$\mathbf{26.6K}_{\pm 15.4K}$	$77.9_{\pm 68.4}$	$85.8_{\pm 80.3}$	$86.1_{\pm 79.5}$	$100.1_{\pm 94.0}$	$\mathbf{22.8}_{\pm 23.5}$
SC-2	$1.2M_{\pm 170.7K}$	$1.1M_{\pm 172.2K}$	$1.2M_{\pm 163.5K}$	$431.9K_{\pm 99.0K}$	$\mathbf{169.1K}_{\pm 34.3K}$	$348.7_{\pm 101.0}$	$1.3K_{\pm 698.2}$	$382.8_{\pm 102.3}$	$338.7_{\pm 181.5}$	$\mathbf{87.3}_{\pm 25.4}$

Performance comparison between the proposed and rule-based strategies with the OptVerse solver

Thank you.