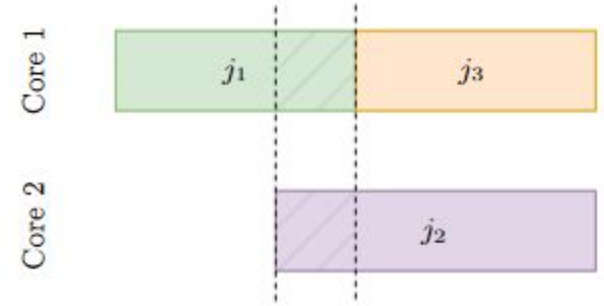
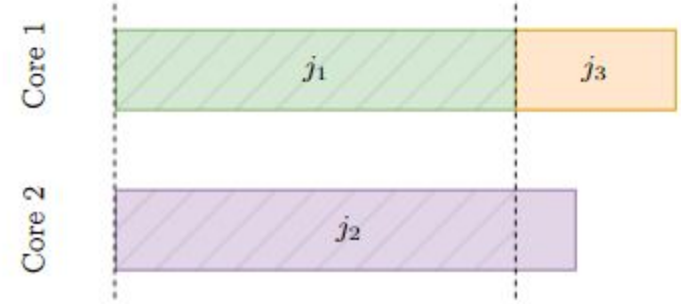
Об оптимизации настройки параметров ПО Gurobi

Устюгов В.Н., Институт математики им. С.Л. Соболева СО РАН (Омский филиал)

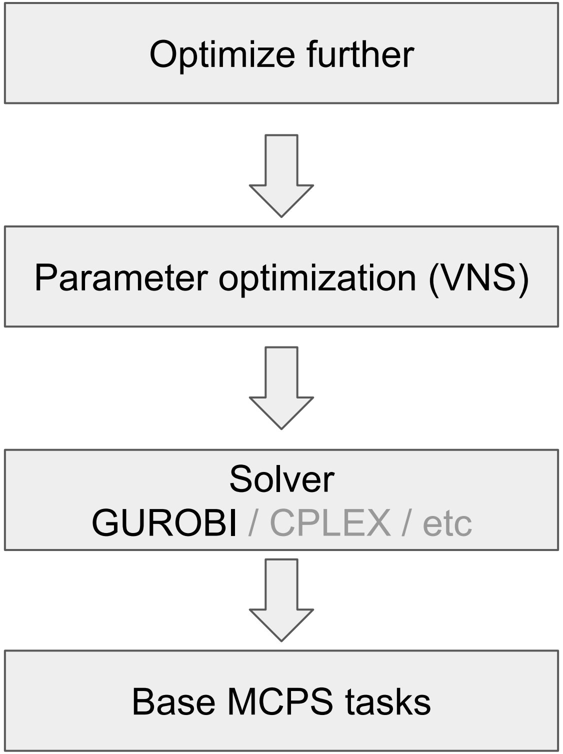
Исследование выполнено за счет гранта Российского научного фонда № 22-71-10015, https://rscf.ru/en/project/22-71-10015/

Базовая задача

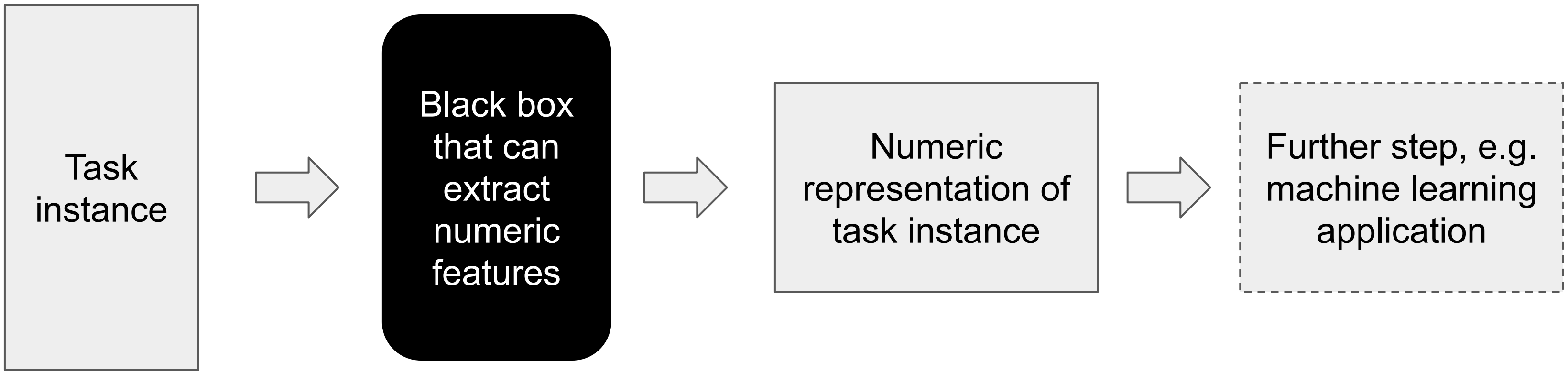




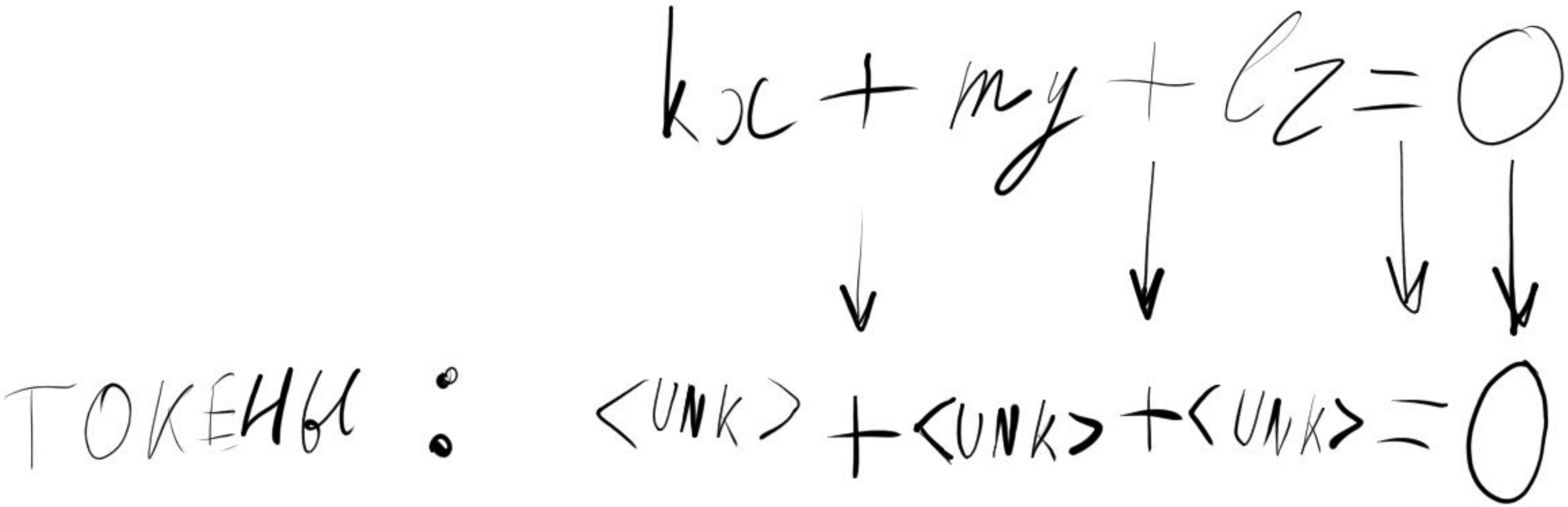
Процесс оптимизации



Процесс оптимизации



Что не так с LLM?



Feature extraction

MIPLIB 2017 – The Mixed Integer Programming Library

In response to the needs of researchers for access to real-world mixed integer programs, Robert E. Bixby, E.A. Boyd, and R.R. Indovina created in 1992 the MIPLIB, an electronically available library of both pure and mixed integer programs. Since its introduction, MIPLIB has become a standard test set used to compare the performance of mixed integer optimizers. Its availability has provided an important stimulus for researchers in this very active area. The library has now been released in its sixth edition as a collaborative effort between Arizona State University, COIN-OR, CPLEX, FICO, Gurobi, MathWorks, MIPCL, MOSEK, NUOPT, SAS, and Zuse Institute Berlin. Like the previous MIPLIB 2010, two main sets have been compiled from the submissions. The **Benchmark Set** contains 240 instances that are solvable by (the union of) today's codes. For practical reasons, the benchmark instances were selected subject to various constraints regarding solvability and numerical stability. The much larger **Collection Set** represents a diverse selection regardless of the above, benchmark-relevant criteria. Download the instance sets as well as supplementary data, run scripts and the solution checker from our Download page.

Feature extraction

4.3 Instance features

Here, we describe the first nine feature groups in Table 4. We use the shorthand *vector statistics* to refer to five values summarizing the entries of a vector $v \in \mathbb{R}^d_{\pm\infty}$. Let $d' = |\{j : |v_j| < \infty\}|$ be the number of finite entries of v, which can be smaller than d in the case of, e.g., bound vectors, and let v' be the restriction of v to its finite entries. We assume without loss of generality that v' is sorted, $v'_1 \le v'_2 \le \cdots \le v'_{d'}$. The five values are

- $\min : v \mapsto v'_1$,
- $\max : v \mapsto v'_{d'}$,
- $\text{mean} : v \mapsto \frac{1}{d'} \sum_{j=1}^{d'} v'_j$,
- $\text{median} : v \mapsto \left(v'_{\lfloor \frac{d'+1}{2} \rfloor} + v'_{\lceil \frac{d'+1}{2} \rceil} \right) /2$, and
- $\text{std} : v \mapsto \sqrt{\frac{1}{d'} \sum_{j=1}^{d'} \left(v'_j - \text{mean}(v') \right)^2}$.

Note that infinite entries can only occur for the variable bound vectors ℓ^x and u^x and the left- and right-hand side vectors ℓ^A, u^A. For a vector v that contains only infinite entries, i.e., for which $d' = 0$, the above vector summaries are not well-defined. If $d' = 0$, the corresponding statistics were set to 0 in the data. Note that even if the

Table 4 Description of instance features used. Set notation is abbreviated, e.g., $\{A \neq 0\}$ denotes $\{(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\} : a_{i,j} \neq 0\}$

Group	Features	Description	Scaling
SIZE	3	Size m,n of matrix, nonzero entries $\lvert\{A \neq 0\}\rvert$	$\log_{10}(x)^2$
VARIABLE TYPES	3	Proportion of binary, integer, and continuous variables $\frac{n_b}{n}, \frac{n_i}{n}, \frac{n_c}{n}$	
OBJECTIVE NONZERO DENSITY	5	Nonzero density of objective function $\frac{\lvert\{c \neq 0\}\rvert}{n}$ both total and by variable type (bin., int., cont.), 0–1 indicator for feasibility problems without objective	
OBJECTIVE COEFFICIENTS	6	Vector statistics and dynamism of c	c normalized by $\lVert c \rVert_\infty$
VARIABLE BOUNDS	12	Finite densities $\frac{\lvert\{\lvert\ell^x\rvert < \infty\}\rvert}{n}, \frac{\lvert\{\lvert u^x\rvert < \infty\}\rvert}{n}$ of bounds, vector statistics of upper bounds u^x and bound ranges $u^x - \ell^x$.	Vector statistics scaled by siglog(x)
MATRIX NONZEROS	6	Vector statistics of nonzero entries $\lvert\{a_i \neq 0\}\rvert$ by row in A, nonzeros per column $\frac{\lvert\{A \neq 0\}\rvert}{n}$	$log_{10}(x)$ for nonzeros per column
MATRIX COEFFICIENTS	19	Vector statistics of the four m-dimensional vectors describing the min, mean, max, and std of the nonzero coefficients $\{a_i \neq 0\}$ in each row	Every a_i normalized by $\lVert a_i \rVert_\infty$
ROW DYNAMISM	5	Vector statistics of row dynamism $\frac{\lVert a_i \rVert_\infty}{\min_j \{\lvert a_{ij}\rvert \neq 0\}}$	$log_{10}(x)$
SIDES	19	Vector statistics of left- and right-hand sides ℓ^A, u^A and concatenated $(\lvert\ell^A\rvert \Vert \lvert u^A\rvert)$, nonzero and finite densities of ℓ^A, u^A	Every a_i normalized by $\lVert a_i \rVert_\infty$
CONSTRAINT CLASSIFICATION	17	Proportion of classes of special linear constraints: singleton, precedence, knapsack, mixed binary (see Sect. 4.4)	
DECOMPOSITION	10	Features describing decomposition D found by GCG with maximum area score (see Sect. 4.5): areascore(D), k, vector statistics (except std) of $\left(\frac{\lvert D^r_1\rvert}{m}, \dots, \frac{\lvert D^r_k\rvert}{m}\right)^\top$ and $\left(\frac{\lvert D^c_1\rvert}{n}, \dots, \frac{\lvert D^c_k\rvert}{n}\right)^\top$. Not available for all instances	

Feature extraction

original formulation has infinite bounds on variables, trivial presolving may often infer finite bounds for those variables.

The *dynamism* of a vector with finite entries is the ratio of the largest and smallest absolute entries, i.e., $\|v\|_\infty / \min\{|v_j| : v_j \neq 0\}$. The dynamism is always at least 1. If the dynamism of any single constraint exceeds 10^6, this is an indication of a numerically difficult formulation. Note that the dynamism is invariant to the normalization procedure. Combining the dynamism of each constraint yields an m-dimensional vector, which can be summarized using vector statistics.

The feature group MATRIX COEFFICIENTS summarizes the nonzero coefficients of the matrix A as follows. First, each row a_i, $1 \leq i \leq m$, of A is normalized by its largest absolute coefficient, such that all coefficients are in the range $[-1, 1]$. The nonzero entries of a_i are then summarized by four of the five vector statistics explained above, namely the min, max, mean, and std. Going through all rows, we obtain four m-dimensional vectors describing the min, max, mean, and std per row. Each of these vectors is then summarized via vector statistics, which yields a total of 20 statistics that summarize the cofficients of A. Examples are the mean minimum coefficient over all m rows, or the standard deviation of all m maximum coefficients, etc. The feature group comprises 19 out of these 20 coefficient statistics, because the maximum over all m maximum coefficients is equal to 1 for every instance in our data set.

For the feature group SIDES, the m-dimensional left- and right-hand side vectors ℓ^A and u^A are summarized individually via vector statistics of all their finite elements. Besides, we compute vector statistics for the finite elements of the concatenated $2m$-dimensional vector $(|\ell^A|, |u^A|)$ that combines the absolute left- and right-hand sides of all rows. Note that the row normalization by the maximum absolute coefficient also affects the row left- and right-hand sides.

For features such as the row or objective dynamism, which may differ by orders of magnitude between instances, we used a logarithmic scaling. While logarithmic scaling is fine for vectors with positive entries, it is not applicable to vectors with potentially negative entries such as the variable lower and upper bound vectors. In those cases, we apply a customized scaling

$$\text{siglog}: \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \text{sig}(x) \log_{10}(|x| + 1)$$

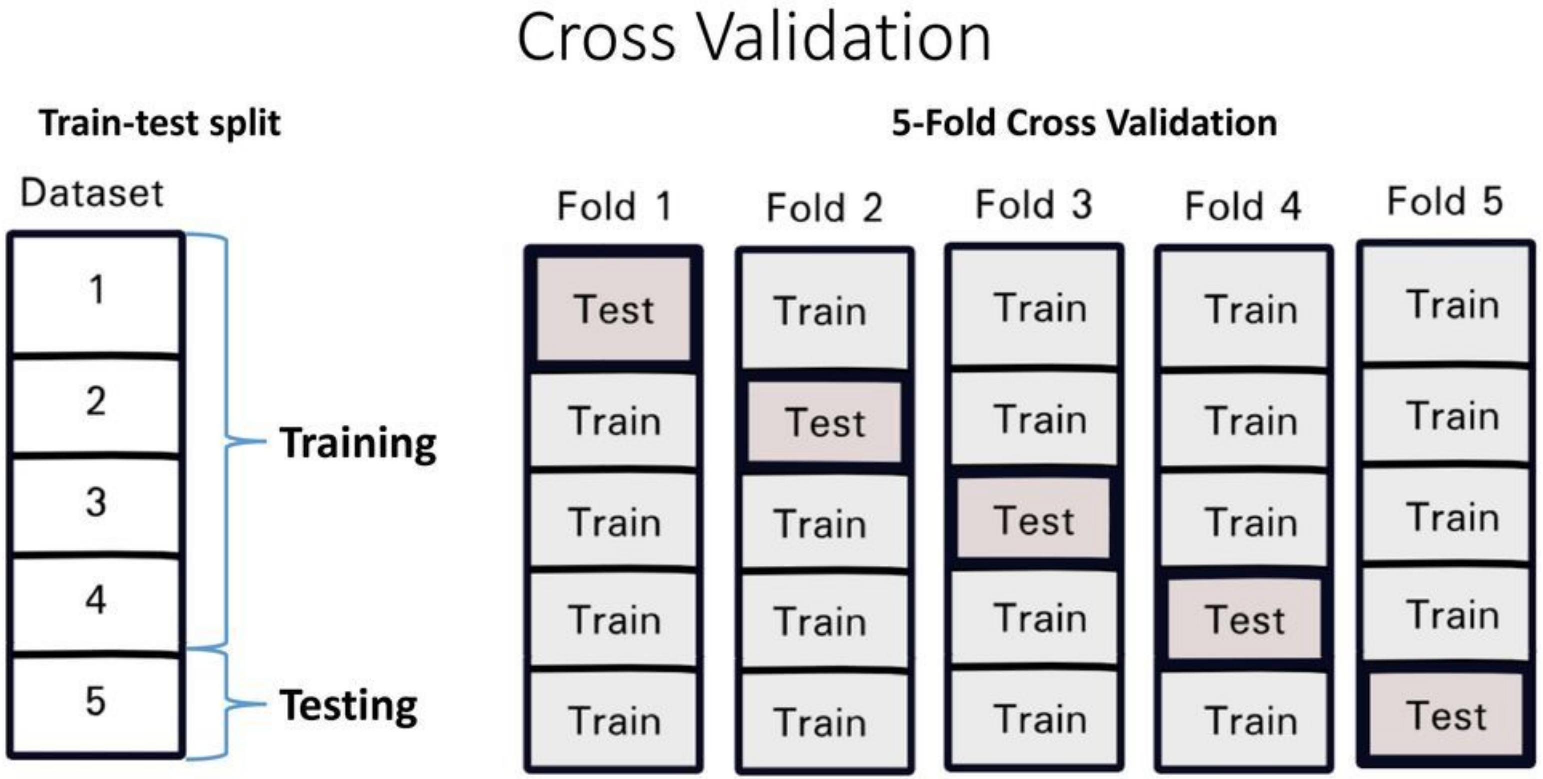
to every entry of the corresponding column in the feature matrix F. The map siglog preserves the sign of each entry.

The collection of the instance features was performed with a small C++ application called the *feature extractor*, which extends SCIP by the necessary functionality needed to report features after trivial presolving and optionally accepts a settings file to modify the default presolving explained in Sect. 4.1. The feature extractor is a modified version of a code used already by [22] and available for download on the MIPLIB 2017 web page (see Sect. 6.3). [5]

[5] The actual computations reported in the following were carried out with five additional, redundant matrix features. Only during the preparation of the manuscript, they were identified to be identical to other features.

Процесс оптимизации

Cross-validation



Результаты численных экспериментов

	Настроенный с VNS оптимум	Предсказанные во время CV
Соотношение среднего времени выполнения в сравнении с запуском на параметрах по-умолчанию	76%	88%
Временные затраты	примерно 40 часов	примерно 3 секунды

Спасибо за внимание!

Литература:

1. MIPLIB 2017: Data-Driven Compilation of the 6th Mixed-Integer Programming Library, Gleixner et al.

Feature extraction

4.2 Canonical form

Because the feature computation can be affected not only by presolving but also by the exact form in which the instance is represented (equality constraints versus inequalities etc.), we transformed all presolved instances into the following canonical form, which is slightly more general than the usual one, prior to feature computation.

Definition 1 A *mixed integer optimization problem* P with input

- $m, n, n_b, n_i, n_c \in \mathbb{N}$, $n = n_b + n_i + n_c$,
- *coefficient matrix* $A \in \mathbb{Q}^{m \times n}$,
- *left-hand and right-hand side vectors* $\ell^A, u^A \in \mathbb{Q}^m_{\pm\infty}$,
- *lower and upper bound vectors* $\ell^x, u^x \in \mathbb{Q}^n_{\pm\infty}$, and
- *objective coefficient vector* $c \in \mathbb{Q}^n$

is defined to be an optimization problem of the form

$$\min \left\{ c^\top x : \ell^A \le Ax \le u^A, \ell^x \le x \le u^x, x \in \{0,1\}^{n_b} \times \mathbb{Z}^{n_i} \times \mathbb{Q}^{n_c} \right\}.$$