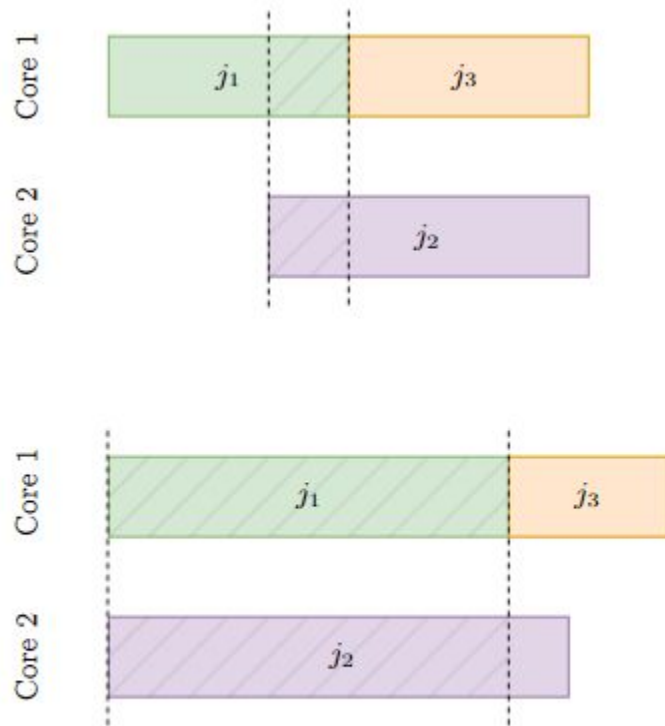


О нелинейных методах машинного обучения в контексте настройки ПО Gurobi

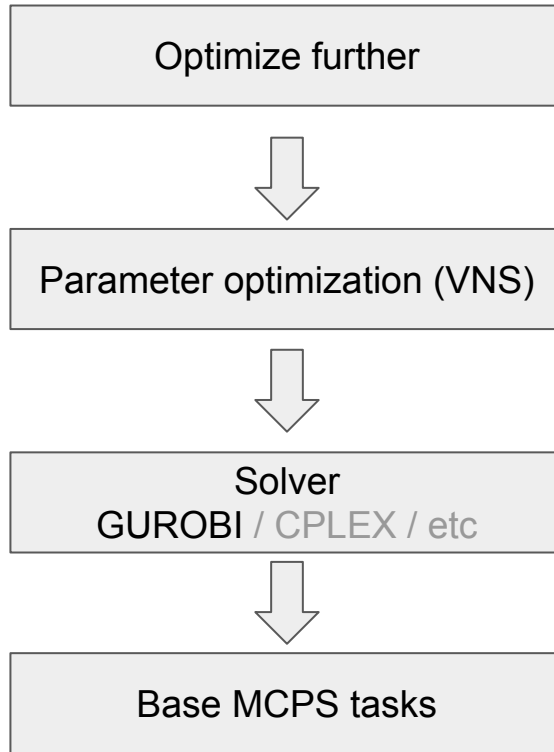
Устюгов В.Н., Институт математики им. С.Л. Соболева СО РАН (Омский филиал)

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

Базовая задача



Процесс оптимизации



Что с данными?

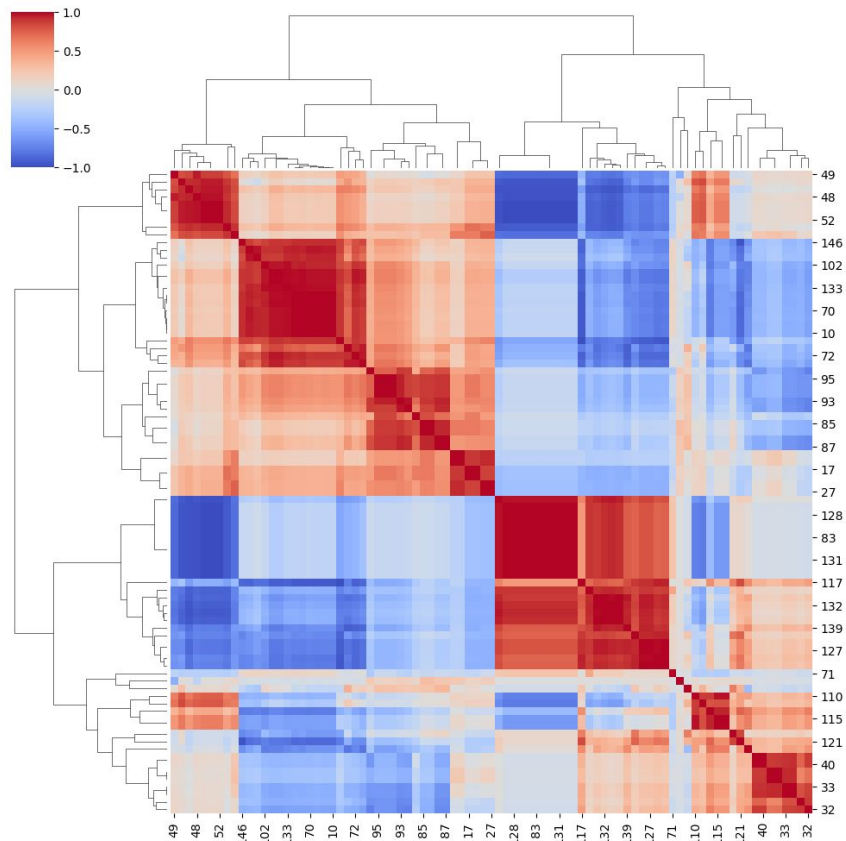
Ограничения:

- Threads строго 8;
- Minrelnodes строго 0.

Параметры выбранные для настройки:

- Gomorypasses;
- Method;
- Mipfocus;
- Presolve.

Данные формата MIPLIB17 и предыдущие результаты

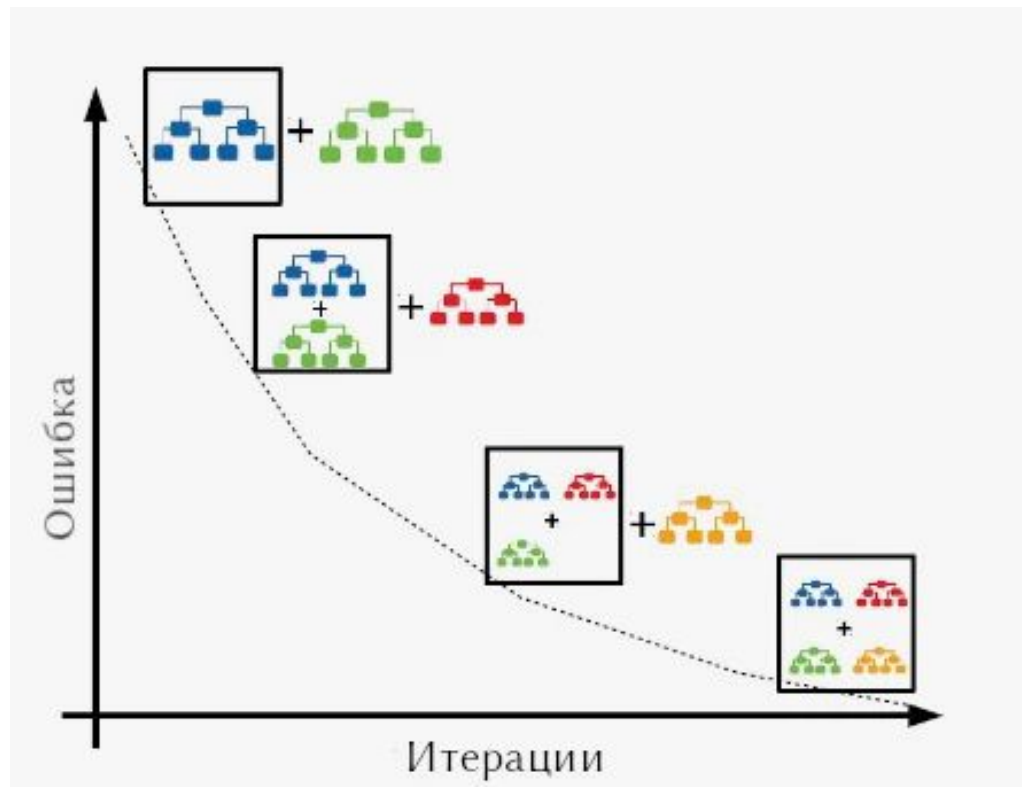


Данные формата MIPLIB17 и предыдущие результаты

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

Градиентный бустинг



Градиентный бустинг

$$\mathcal{L}(y, x) = \frac{1}{2} \sum_{i=1}^N (y_i - a(x_i))^2 \rightarrow \min$$

Отметим теперь важное свойство функции потерь в рассмотренном выше примере с регрессией. Для этого посчитаем производную функции потерь по предсказанию

$z = a_k(x_i)$ модели для i -го объекта:

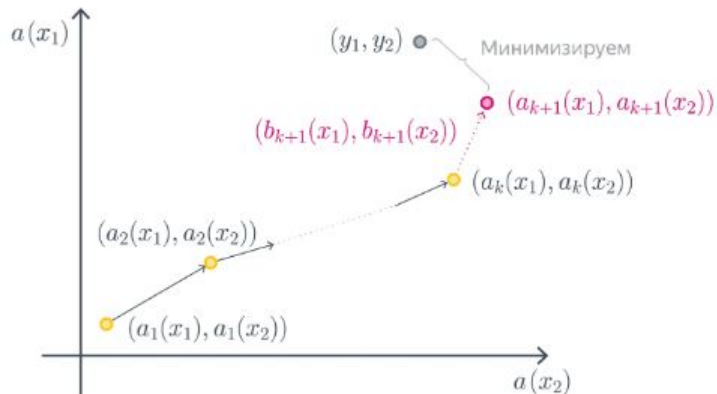
$$\left. \frac{\partial \mathcal{L}(y_i, z)}{\partial z} \right|_{z=a_k(x_i)} = \left. \frac{\partial}{\partial z} \frac{1}{2} (y_i - z)^2 \right|_{z=a_k(x_i)} = a_k(x_i) - y_i$$

Видим, что разность, на которую обучается k -й алгоритм, выражается через производную:

$$s_i^k = y_i - a_k(x_i) = - \left. \frac{\partial \mathcal{L}(y_i, z)}{\partial z} \right|_{z=a_k(x_i)}$$

Градиентный бустинг

Вспомните аналогию с гольфистом: обучение композиции можно представить как перемещение предсказания из точки $(a_k(x_1), a_k(x_2), \dots, a_k(x_N))$ в точку $(a_{k+1}(x_1), a_{k+1}(x_2), \dots, a_{k+1}(x_N))$. В конечном итоге мы ожидаем, что точка $(a_K(x_1), a_K(x_2), \dots, a_K(x_N))$ будет располагаться как можно ближе к точке с истинными значениями (y_1, y_2, \dots, y_N) .



Чем хорош градиентный бустинг?

- Не боится разных масштабов в данных, пропусков, мультиколлинеарности, нелинейных и категориальных данных;
- Похож на нейросеть возможностями настройки и методологией обучения;

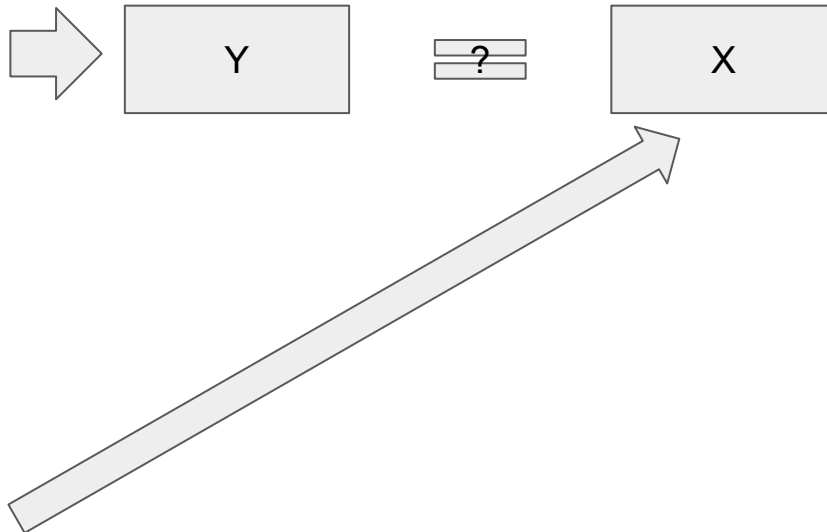
Постановка эксперимента

VNS поверх Gurobi
на задачах составления
расписания

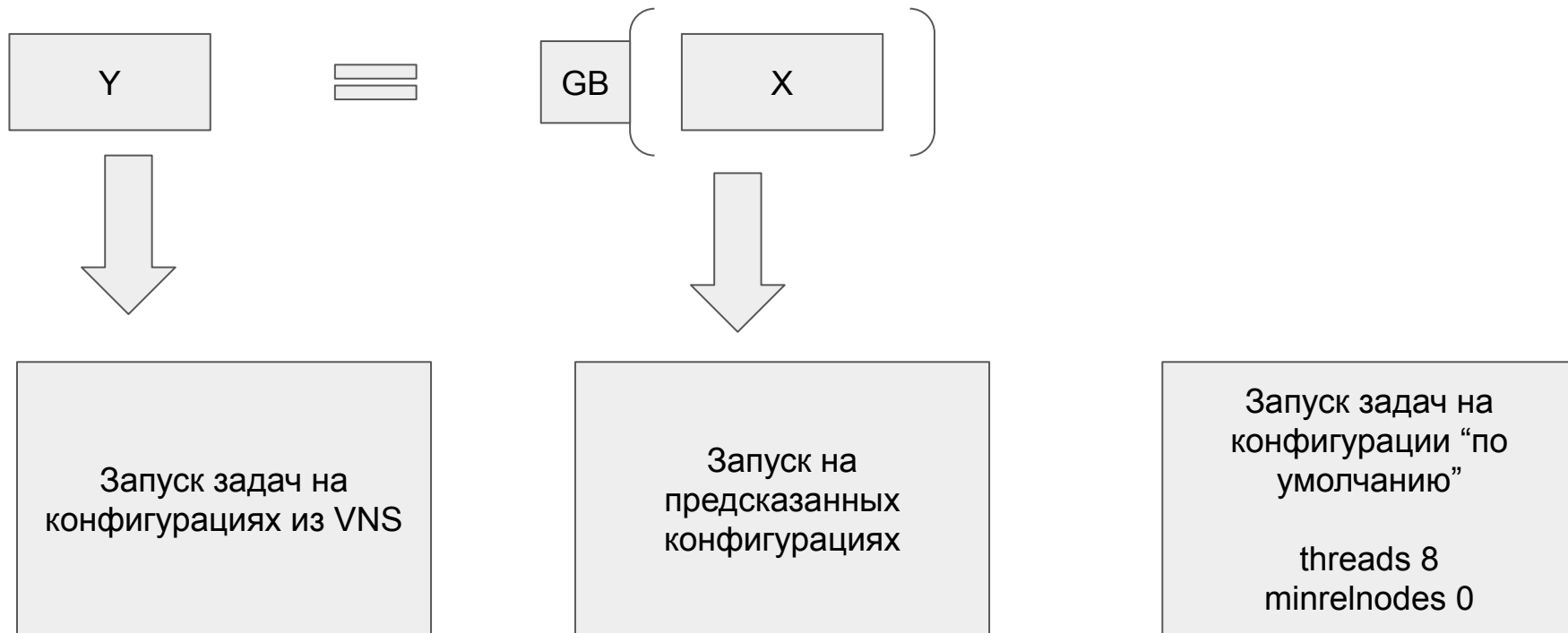
для каждой задачи своя
конфигурация Gurobi

Feature extraction
методом MIPLIB

Удобно подходит как
входная информация о
задаче



Постановка эксперимента



Тест Вилкоксона

Легенда:

default - выборка времён запусков, полученная на конфигурации по-умолчанию;

individual - выборка времён запусков, полученная на найденных VNS конфигурациях

predicted - выборка времён запусков, полученная с помощью градиентного бустинга в эксперименте с методикой leave-one-out

Основная гипотеза	default < predicted	predicted < individual
p-value	2.74e-29	0.67
Комментарий	Гипотеза отвергается, значит predicted статистически значимо меньше, чем default	Гипотеза не отвергается. В обратную сторону, кстати, тоже не отвергается, хотя по сырым данным похоже, что individual лучше predicted

Спасибо за внимание!

Литература:

1. MIPLIB 2017: Data-Driven Compilation of the 6th Mixed-Integer Programming Library, Gleixner et al.
2. CatBoost Web Documentation, <https://catboost.ai/docs/en/>

Cross-validation

Cross Validation

