



Санкт-Петербургский
Государственный
Политехнический
Университет

Институт прикладной
математики и механики

КАФЕДРА

ТЕЛЕМАТИКА

Управление научными проектами (Методы исследовательской работы)

Исследования как процесс получения новых знаний : первые принципы, энергетика обучения и процессы вычисления (занятие 6)

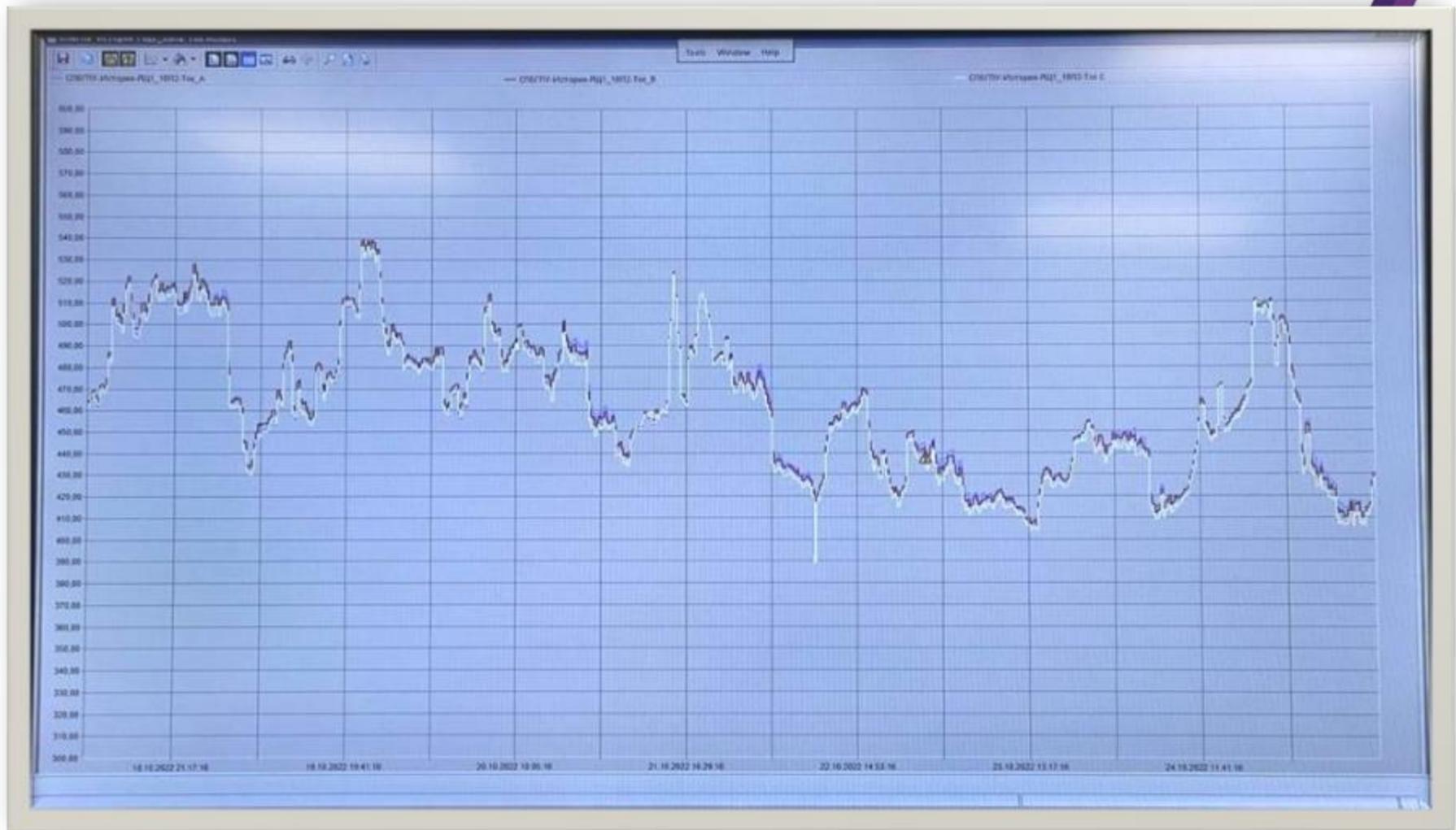
10, 17, 24 октября
2023 г.



СОДЕРЖАНИЕ

- «Первые принципы» науки и ошибки мышления
<https://liferhacker.ru/30-oshibok-myshleniya/>
- **Пространство возможностей** увеличения производительности
- Концепция «обучение + программирование» Что конкретно должно «обучается» в СКЦ?
- **Однако «все уже было».** Формализация задачи обучения СК

ЭНЕРГЕТИКА ПРОЦЕССОВ ВЫЧИСЛЕНИЙ: О ЧЕМ ГОВОРИТ ТАКОЙ ГРАФИК ПОТРЕБЛЕНИЯ ЭЛ. ТОКА ВЫЧИСЛИТЕЛЬНЫМИ УЗЛАМИ СК



ИНФОРМАЦИОННОЕ СОДЕРЖАНИЕ СЛ. ВЕЛИЧИНЫ

Кластер это многокомпонентная система, которая находится все время в неравновесном состоянии и характеризуется:

- - состоянием отдельных своих подсистем и прикладных задач пользователей
- - взаимным влиянием, которое осуществляется через диспетчер, различных подсистем кластера на эффективность работы кластера в целом,

Роль диспетчера в том, чтобы используя информацию (что такое информация ?) о состоянии узлов кластера и потребности обрабатываемых заданий, распределить ресурсы так, чтобы задания были выполнены с кодом завершения «0» то есть «успешно».

- В кластере происходит постоянное «движение» ресурсов, которые занимают в момент времени, который назначен диспетчером, освобождаются в случайный момент времени. Вариантов занять определенное количество узлов имеющимся множеством задач существует много

ПЕРВЫЕ ПРИНЦИПЫ НАУКИ- ЭЛЕМЕНТЫ ИСТИННОГО ЗНАНИЯ *THE BUILDING BLOCKS OF TRUE KNOWLEDGE*

Большинство людей учатся не через понимание,
а зубрежкой, поэтому их знания так хрупки!

Р. Фейнман

- Считается, что : истинные научные знания (а не аналогии или корреляции) могут быть выражены (выведены) через «первые принципы», которые применимы ко всем объектам Природы
- На «первых принципах» основан процесс управления научными проектами, позволяя «продраться сквозь туман ложных убеждений и неэффективных аналогий»

«ПЕРВЫЕ ПРИНЦИПЫ»

“Хотя всякое наше познание и начинается с опыта, отсюда вовсе не следует, что оно целиком происходит из опыта”.

И. Кант

(кн. Критика чистого разума)

- это базовое положение или допущение, которое не может быть выведено ни из какого другого положения или допущения.
- Аристотель:
«В каждом систематическом изыскании (греч. methodos), где существуют первые принципы, любые знание и науки являются результатом познания этих принципов.
- Познать «что-либо», это значит узнать о том, какие причины и принципы объединяют различные элементы в то, что воспринимаемое как единое целое
- "Мышление на основе первых принципов" состоит из разложения вещей до фундаментальных элементов (не делимых далее), а затем — реконструирование их воедино, создавая нечто новое и лучшее.

ПРИМЕР

- Энергия способна менять свою форму. Кинетическая, потенциальная, химическая, электрическая, ядерная и **тепловая** – это лишь некоторые формы энергии. Энергия постоянно переходит из одной формы в другую, но **одно лишь неизменно**: полная сумма значений энергии по всем ее формам никогда не меняется.
- **Фактология**: единственным результатом любой совокупности процессов **не может быть** только **превращение теплоты в полезную (?) работу**.
- **Тепло** – это форма энергии случайного хаотического движения, а энтропия – это количество скрытой **микроскопической** информации.
- **«Первый принцип»**: горячие тела остывают, холодные нагреваются, энергия рассеивается, но полная сумма всех форм энергии остается постоянной.

«ПЕРВЫЕ ПРИНЦИПЫ» VS АНАЛОГИИ

Я считаю, что важно мыслить, отталкиваясь от первых принципов, а не от аналогий... Первый принцип — это взгляд на мир с точки зрения физики — сведение всех проблем к фундаментальным истинам

И. Маск

Первые принципы надо применять:

- Когда ищется решение сложной проблемы.
- Что-то делается впервые.
- Необходимо разобраться в трудной теме.

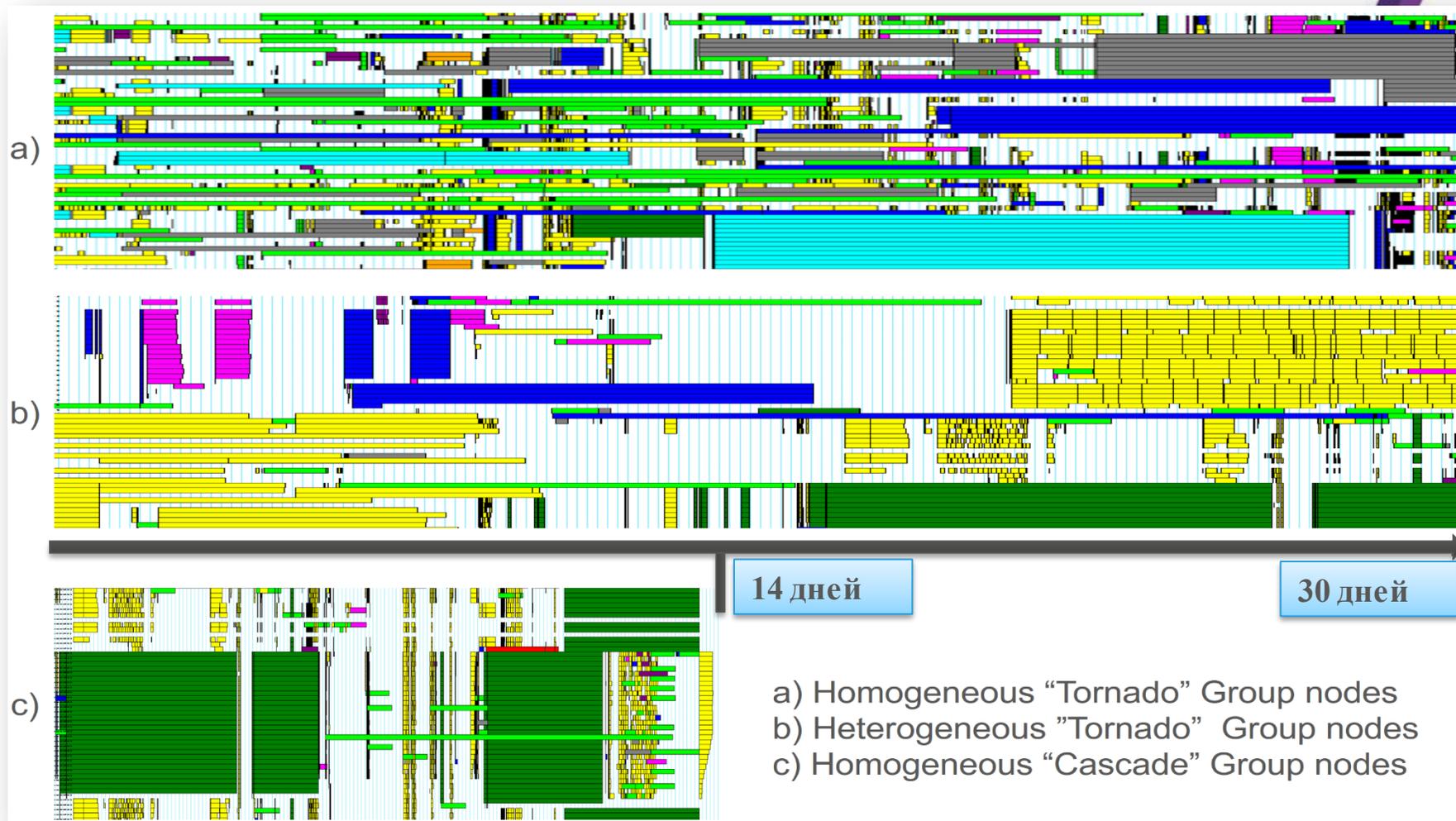
Для того, чтобы применять первые принципы **нужно научиться задавать правильные вопросы!** А ответы на вопросы выражать в фундаментальных понятиях.



ПРЕОБРАЗОВАНИЕ ОПЫТА В НАУЧНЫЕ ЗНАНИЯ – СОВОКУПНОСТЬ БАЗОВЫХ (НЕДЕЛИМЫХ) ПОНЯТИЙ



ПРОСТРАНСТВО ВОЗМОЖНОСТЕЙ УВЕЛИЧЕНИЯ ПРОИЗВОДИТЕЛЬНОСТИ СК: КАК ОБЕСПЕЧИТЬ ЭФФЕКТИВНУЮ ЗАГРУЗКУ УЗЛОВ КЛАСТЕРА



ТОЧНОСТЬ РАСПРЕДЕЛЕНИЯ РЕСУРСОВ = ЗНАНИЕ ТОЧНОГО ВРЕМЕНИ
ЗАВЕРШЕНИЯ ПРИКЛАДНЫХ ЗАДАНИЙ

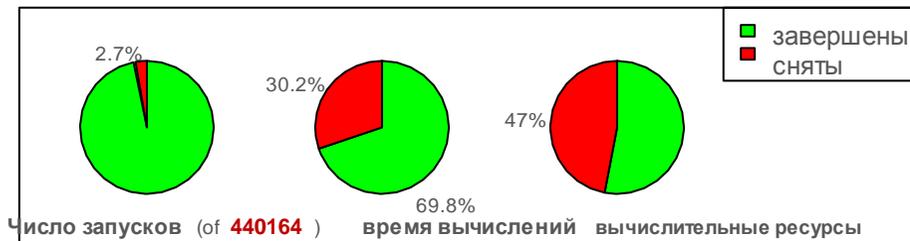


Недоиспользованные ресурсы

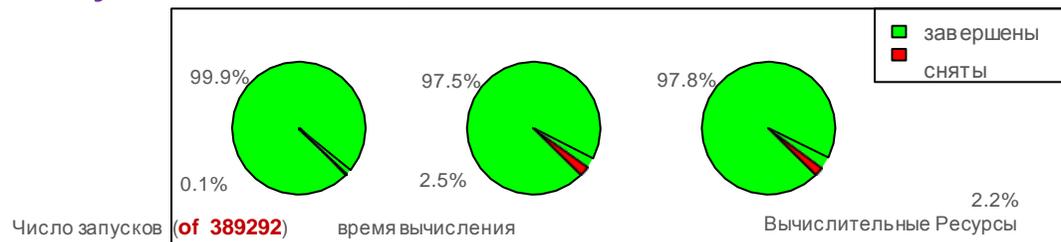


«ПРОСТРАНСТВО ВОЗМОЖНОСТЕЙ» ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ СОСТОИТ ИЗ ЗНАНИЯ ПРИЧИН ВОЗНИКНОВЕНИЯ ПРОБЛЕМ

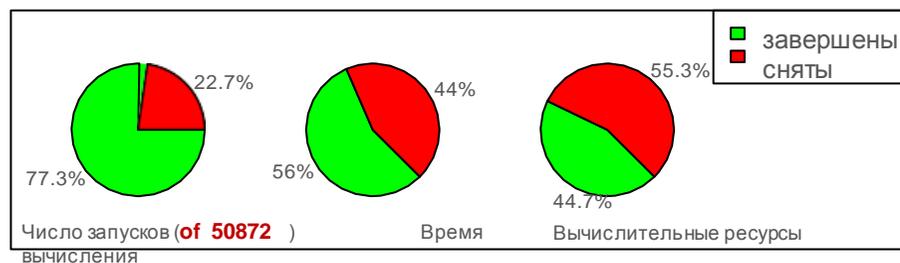
Общая характеристика эффективности для всех видов запусков заданий



Автоматический запуск «известных» заданий ГПН



Запуск «персональных заданий» в ручном режиме



Вывод:

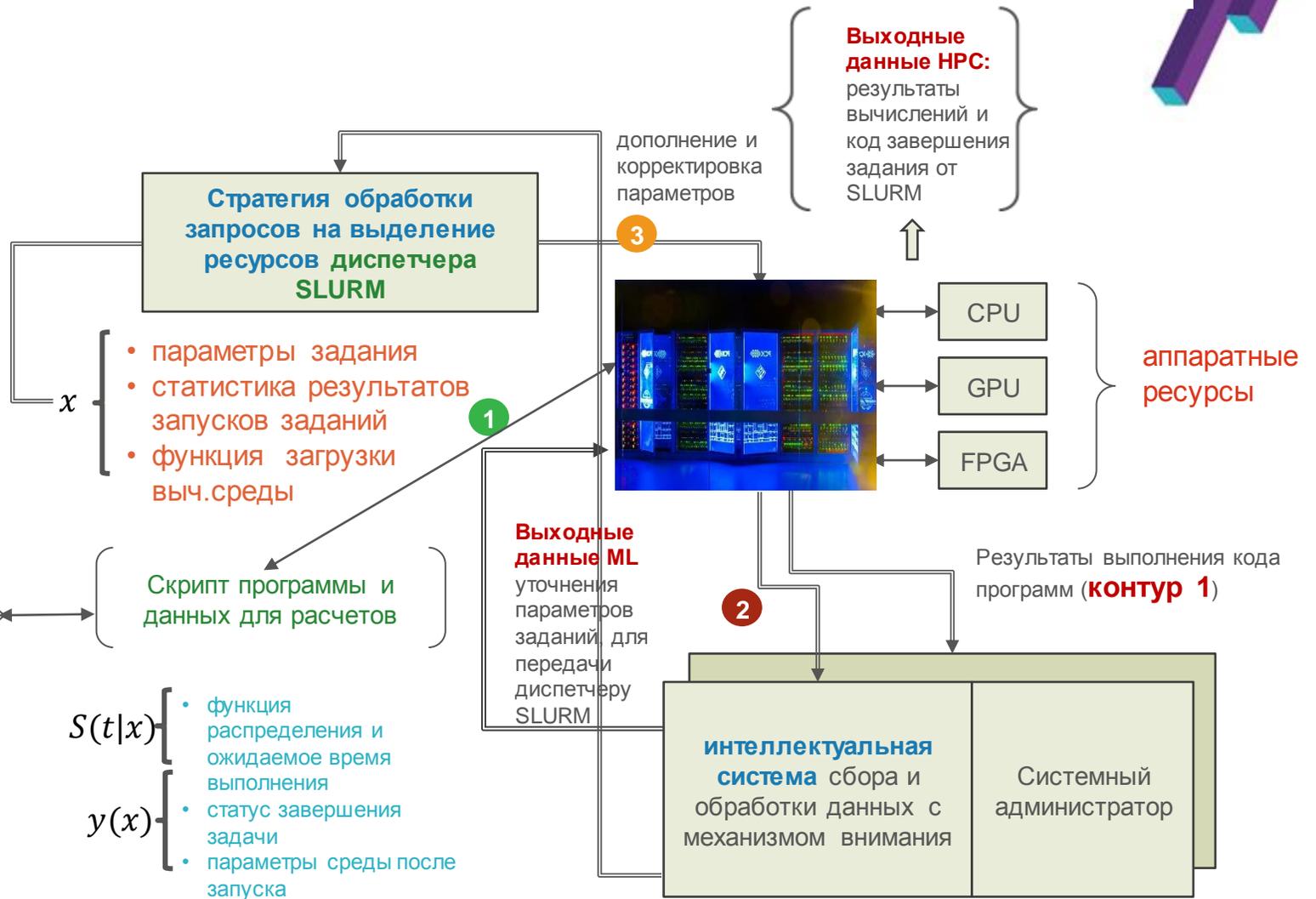
При использовании заданий, параметры которых точно известны, процессы вычислений можно автоматизировать, повышая реальную производительность СК

ПОИСК ЗНАНИЙ И ВОЗМОЖНОСТИ ОБУЧЕНИЯ

Входные данные

- параметры задания,
- модель пользователя
- характеристики среды исполнения

]





ЧЕМУ И ЧТО КОНКРЕТНО ДОЛЖНО «ОБУЧАЕТСЯ» СКЦ:

- Система управления заданиями, используемая для повышения **реальной производительности (число успешно решенных задач)** за счет распределения имеющихся программно-аппаратных ресурсов между задачами пользователей и вычисления (настройки) **функции полезности диспетчера**,
 - которая характеризует **вероятность (возможность) успешного завершения** прикладных вычислительных процессов
- **Цель обучения:** настройка системы управления так, чтобы вероятность успешного завершения была **наибольшей**, а число неиспользуемых узлов кластера **наименьшим**
- **Концепция рефлексивного программирования** - In computer science, reflective programming or reflection is the ability of a process to examine, introspect, and modify its own structure and behavior.
- <https://habr.com/ru/articles/598981/>



ФОРМУЛИРОВКА НОВЫХ ЗНАНИЙ ТРЕБУЕТ НОВЫХ АБСТРАКЦИЙ: РЕАЛЬНАЯ ПРОИЗВОДИТЕЛЬНОСТЬ И «ПОЛЕЗНОСТЬ» СК

Если базироваться на энергетической оценке производительности СК, то объем **«полезной работы»** - суть **«функция»**, которая характеризует долю тех узлов кластера, которые реально используются для выполнения задач пользователей. Для это

- могут использоваться **методы «машинного обучения»** системы распределения имеющихся программно-аппаратных ресурсов (узлов) кластера между прикладными задачами, за счет
 - использования наиболее вероятной оценки времени исполнения заданий и передачи этой оценки диспетчеру для планирования загрузки кластера



ФУНДАМЕНТАЛЬНАЯ ПРОБЛЕМА : ГДЕ ВЗЯТЬ ОБУЧАЮЩУЮ ВЫБОРКУ

Требуется оценки параметров прикладных заданий и динамически изменяющиеся характеристики СК, которые используются системой планирования slurm и непосредственно влияют на реальную производительность работы СК

Для достижения цели используется цензурированная выборка данных, включая

- $x_i \in \mathcal{X}$ – входной вектор параметров задания, задаваемых непосредственно самим пользователем
- $t_i \in [0, +\infty)$ – время до события (завершения задачи)
- $\delta_i \in \{0, 1\}$ – индикатор цензурирования (1, если задача завершена до истечения времени, 0 иначе)
- $y_i \in \mathcal{Y}$ – вектор параметров задачи, не известный на момент запуска (например, факторы изменения состояния системы slurm), в который могут входить индикаторы ошибки выполнения задачи (0 – успешное завершение, 1 – завершение с ошибкой)



ФОРМАЛИЗАЦИЯ ТРЕБОВАНИЙ К ЗАДАЧЕ ОБУЧЕНИЯ:

Построить модель оценки параметров прикладного задания и функции полезности с использованием механизмов внимания в условиях цензурирования данных, а именно:

- ожидаемое время решения $\mathbb{E}[T|X = x]$
- функции выживаемости прикладной задачи в среде СК, а именно $S(t, X)$
- вектора параметров \hat{y} выполнения задачи, компонентами которого являются вероятность успешного выполнения задания $P(X=x)$
- **функция полезности ...**

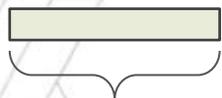
с целью использования этих параметров **в качестве входных данных системы планирования** ресурсов СКЦ, построенной на базе

- стандартного диспетчера **slurm** - *Simple Linux Utility for Resource Management*
- или
- нового планировщика **mlurm** – *Machin Learning Utility for Resource Management*



Надо понимать, что «ВСЕ УЖЕ БЫЛО» : ПОИСК АНАЛОГИЙ ДЛЯ РЕШАЕМОЙ ЗАДАЧИ

Фрагментация дискового пространства для размещения файлов известной длины

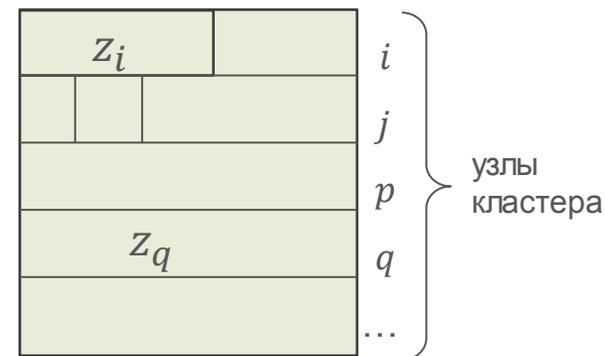
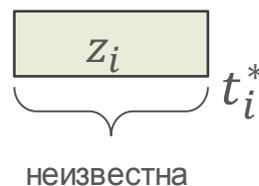


l байт
 l известна



Особенность решаемой задачи

- размещение прикладной задачи z_i в 2D пространстве (номер узла, время занятия узла прикладной задачей) узлов кластера, когда точное время t_i^* не известно, а задается оценка \hat{t}_i
- фактическое время исполнения задачи точно не известно
- состояние кластера – число свободных узлов – меняется
- если $t_i^* > \hat{t}_i$ - задача снимается.

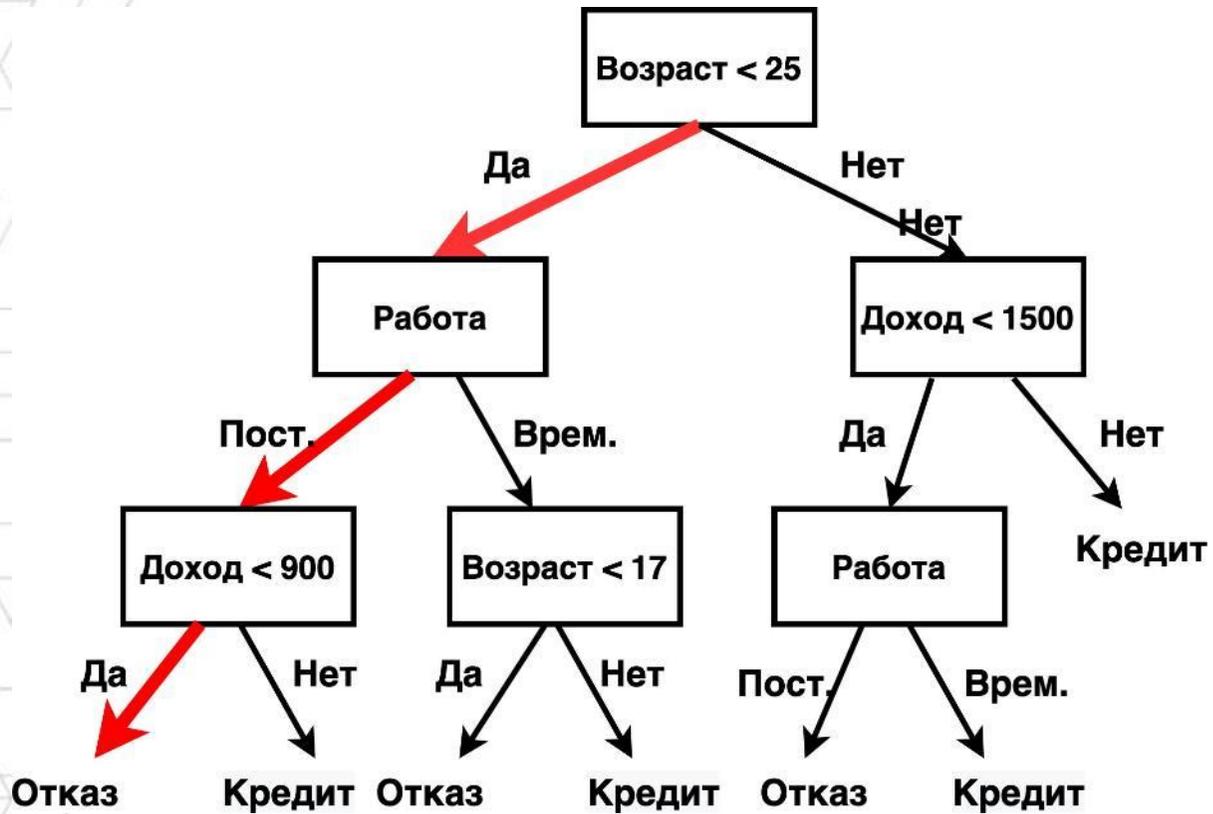




ВЫБОР МОДЕЛИ ОБУЧЕНИЯ : ДЕРЕВЬЯ РЕШЕНИЙ, ПОЧЕМУ ЭТО ЛУЧШЕ, ЧЕМ «ИСКУССТВЕННАЯ НЕЙРОННАЯ СЕТЬ»?

Пояснительный пример:

Имеется количественный «порог» обоснования сделанного выбора



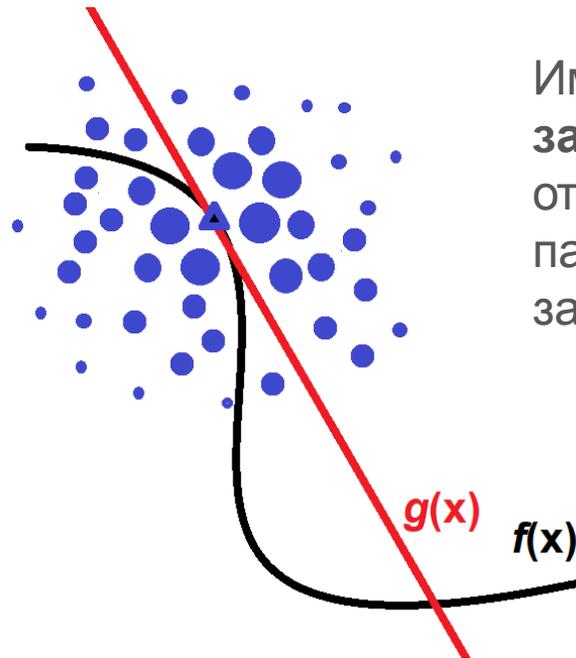
Результат решения можно не только получить, но и объяснить



ВЫБОР МОДЕЛИ ОБЪЯСНЕНИЯ НА ОСНОВЕ РЕГРЕССИИ, ПОЧЕМУ ?

Линейная регрессия: $g(x) = a_1x_1 + a_2x_2 + \dots + a_mx_m$

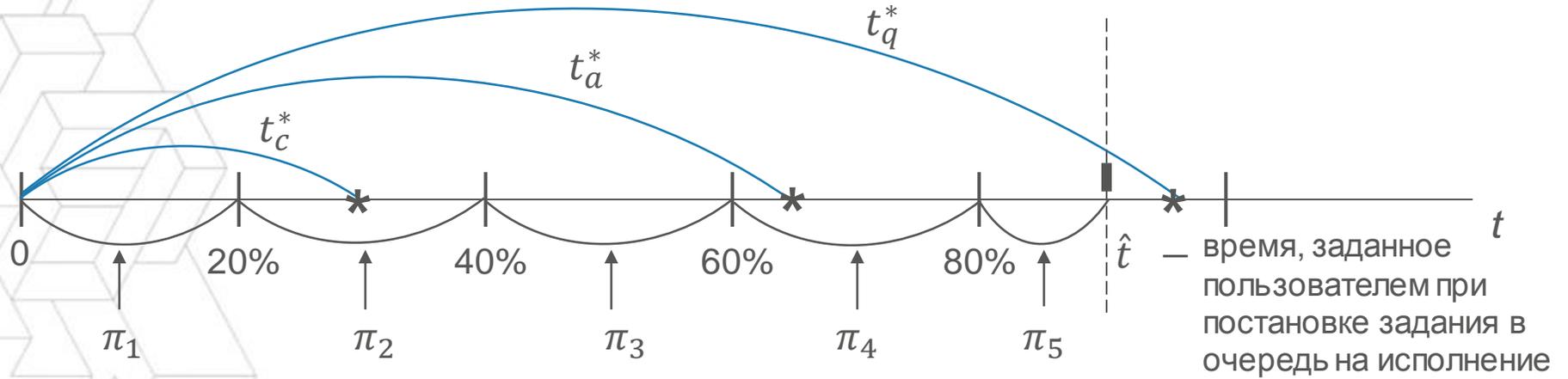
GAM: $g(x) = g_1(x_1) + g_2(x_2) + \dots + g_m(x_m)$



Имеется **линейная зависимость** от локально измеряемых параметров или линейная зависимость от этих функций



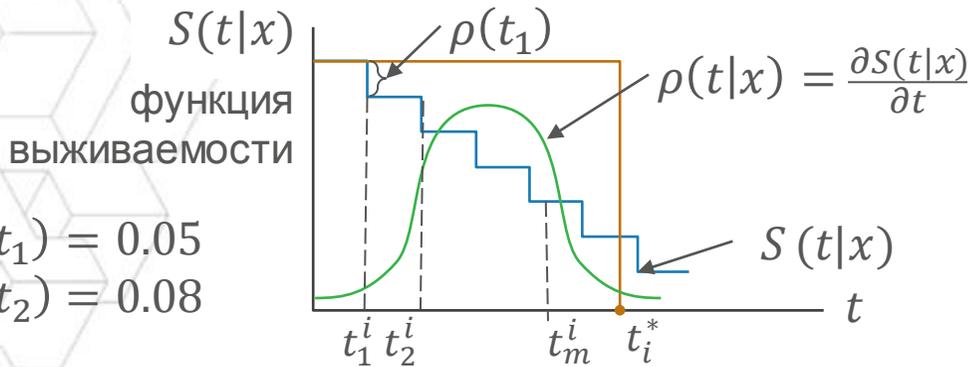
ОСОБЕННОСТЬ ЗАДАЧИ ОБУЧЕНИЯ ДИСПЕТЧЕРА – КТО «ВИНОВАТ» И КОМУ МОЖНО «ДОВЕРЯТЬ»



π_i - доля задач, которые попали в i -й интервал (оценка вероятности того, что)

γ_1 γ_2 ... γ_m

задача q превысила время \hat{t}_q произошло деление на 0 заикливание задачи



$\rho(t_1) = 0.05$
 $\rho(t_2) = 0.08$

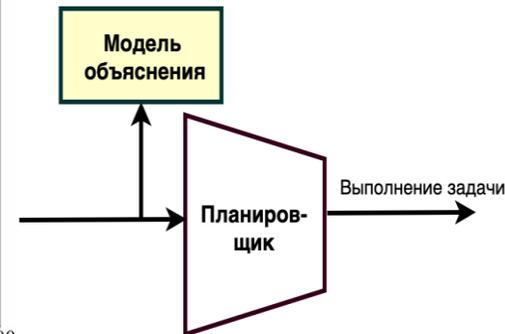
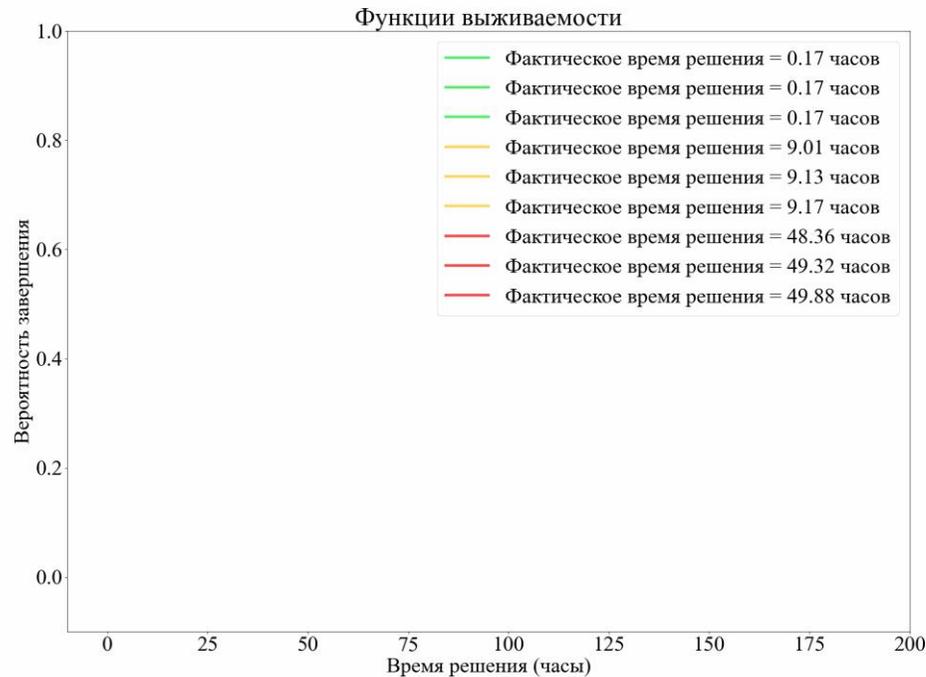
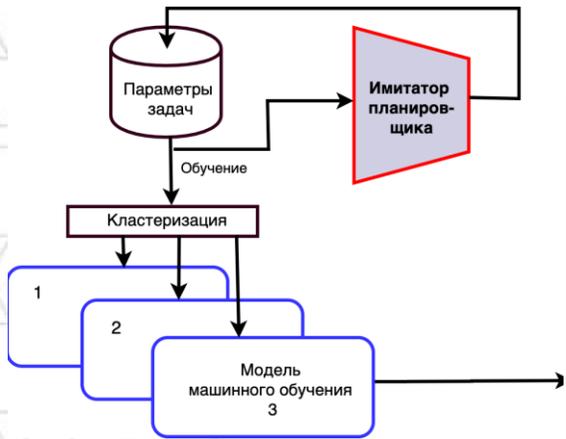
Функция полезности диспетчера:

$$IU = \int_0^{t_{max}} u(t) \cdot \rho(t|x) dt$$

Известно: $u(t)$ – доля занятых узлов кластера
 $S(t|x)$ – функция выживаемости



НОВАЯ АБСТРАКЦИЯ «ФУНКЦИЯ ВЫЖИВАЕМОСТИ» ПРИКЛАДНЫХ ЗАДАЧ В ВЫЧИСЛИТЕЛЬНОЙ СРЕДЕ СУПЕРКОМПЬЮТЕРА $\rho(t|x)dt$



вектор x :



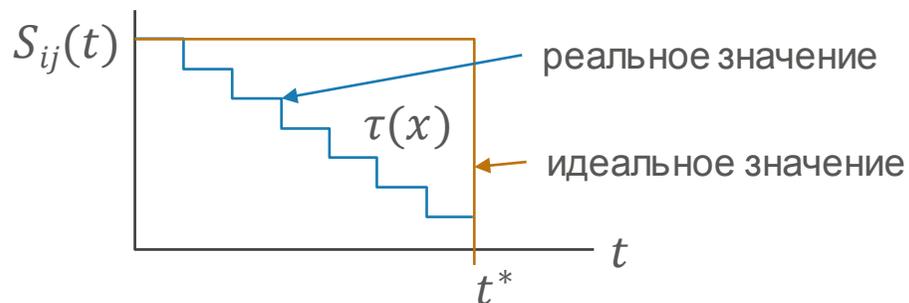
$$(\pi_{1i}, \dots, \pi_{ni}, \rho_{1i}, \dots, \rho_{jp})$$



НОВАЯ АБСТРАКЦИЯ - ФУНКЦИИ ПОЛЕЗНОСТИ ДИСПЕТЧЕРА СК

1. Оцениваем «персональный» эффект от использования функции «выживаемости» прикладной задачи

$S(t|x)$



«персональный» — значит покажем конкретному времени

2. Строим «функцию полезности» интеллектуального диспетчера всего СК

характеризующая «средний» эффект повышения загрузки СК при условии повышения вероятности успешного завершения заданий

в общем случае

$$IU = \int_0^{t_{max}} u(t) \frac{\partial S(t)}{\partial \tau} dt = \int_0^{t_{max}} u(t) \rho(t) dt$$

в нашем случае

$$IU = \sum_{i-\text{пользователь}}^N \sum_{j-\text{задача}}^M \int_0^{t_{max}(i,j)} u_{ij}(t) \rho_{ij}(t|x_{ij}) dt$$

КАК ПОНИМАТЬ «МЕХАНИЗМ ВНИМАНИЯ» В ЗАДАЧАХ МАШИННОГО ОБУЧЕНИЯ

$$y = \sum_{i=1}^N \alpha(q, k_i) v_i$$

$$\alpha(q, k_i) = \operatorname{softmax}_i \left(\operatorname{score}(q, k_j) \right)_{j=1}^N$$

q – query -

k_i – key -

v_i – value -

запрос

ключ

значение

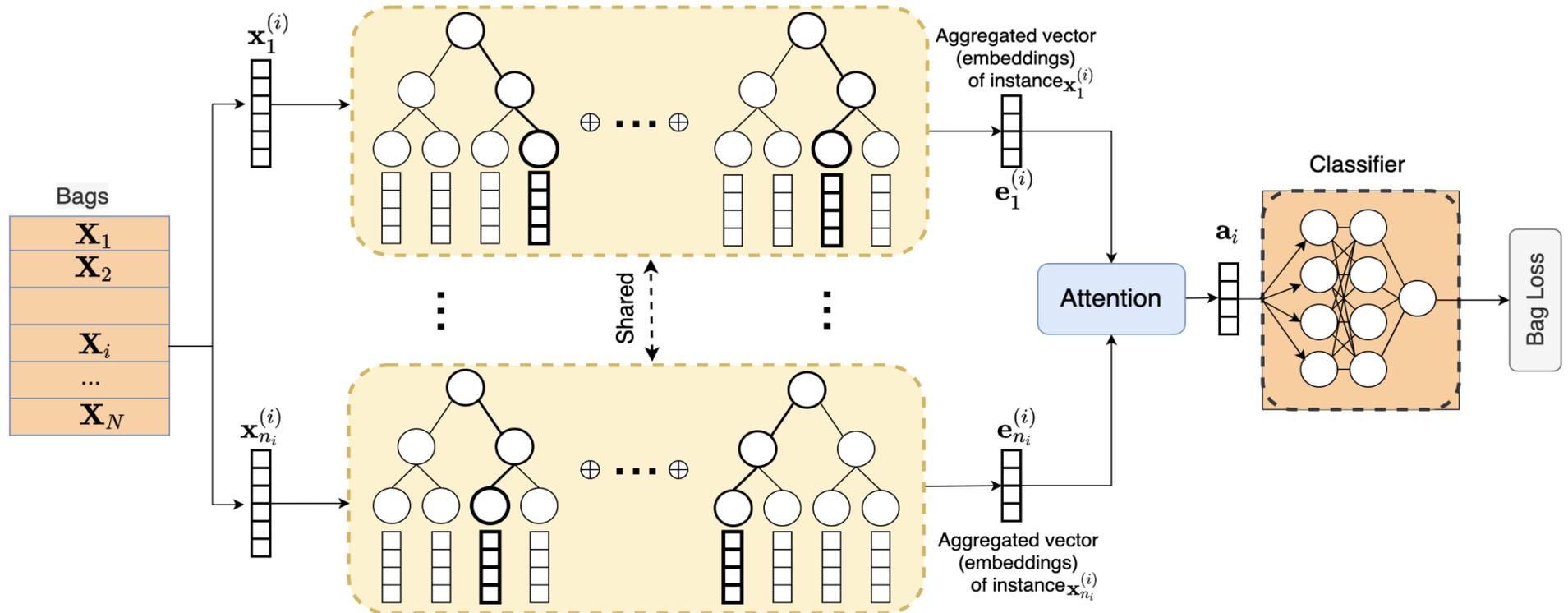
Внимание это - преобразование «ядра» модели $\alpha(q, k_i)$ в «вероятность» некоторого действия

- Такое преобразование характерно для задач обработки сложных последовательностей, например предложений естественного языка
- Позволяет работать с произвольным числом пар (ключ, значение)
- Чаще всего функция $\operatorname{score}(q, k_i) \propto q^T k_i$, то есть линейная
 - либо $(W_Q q)^T (W_K k_i) = q^T W_Q^T W_K k_i = q^T W k_i$
 но допустимо множество и других вариантов

МОДЕЛЬ МНОГОВАРИАНТНОГО ОБУЧЕНИЯ

Предложенная модель может быть эффективно обучена для решения групп задач

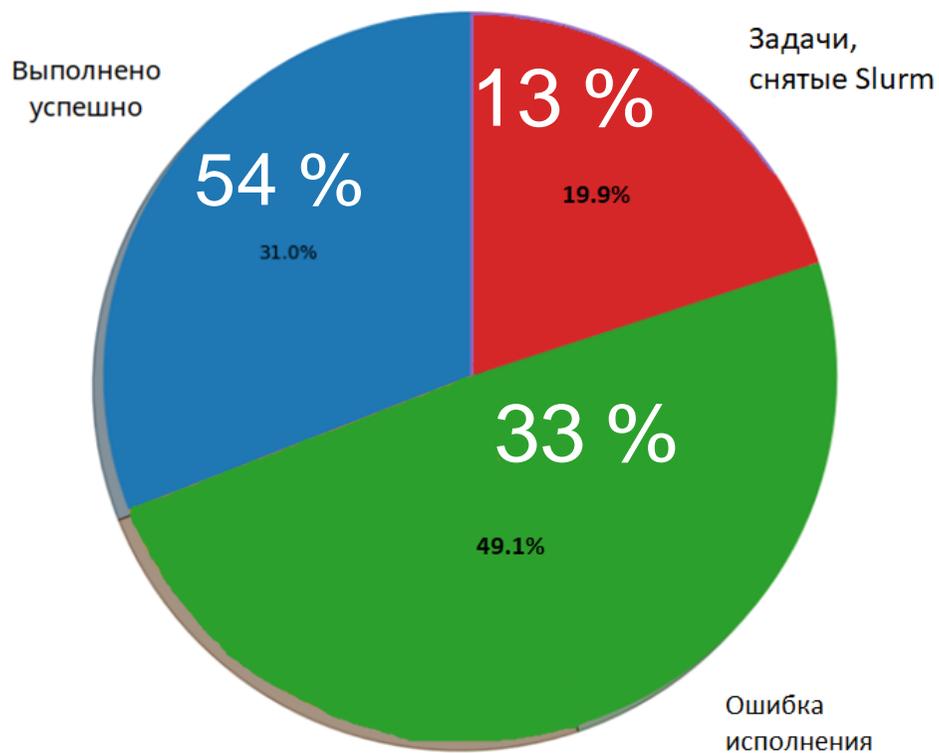
$$D_1 = \{(x_i^1, y_i^1)\}_{i=1}^{N_1}, \dots, D_K = \{(x_i^K, y_i^K)\}_{i=1}^{N_K}$$



Multiple Instance Learning with Trainable Decision Tree Ensembles,

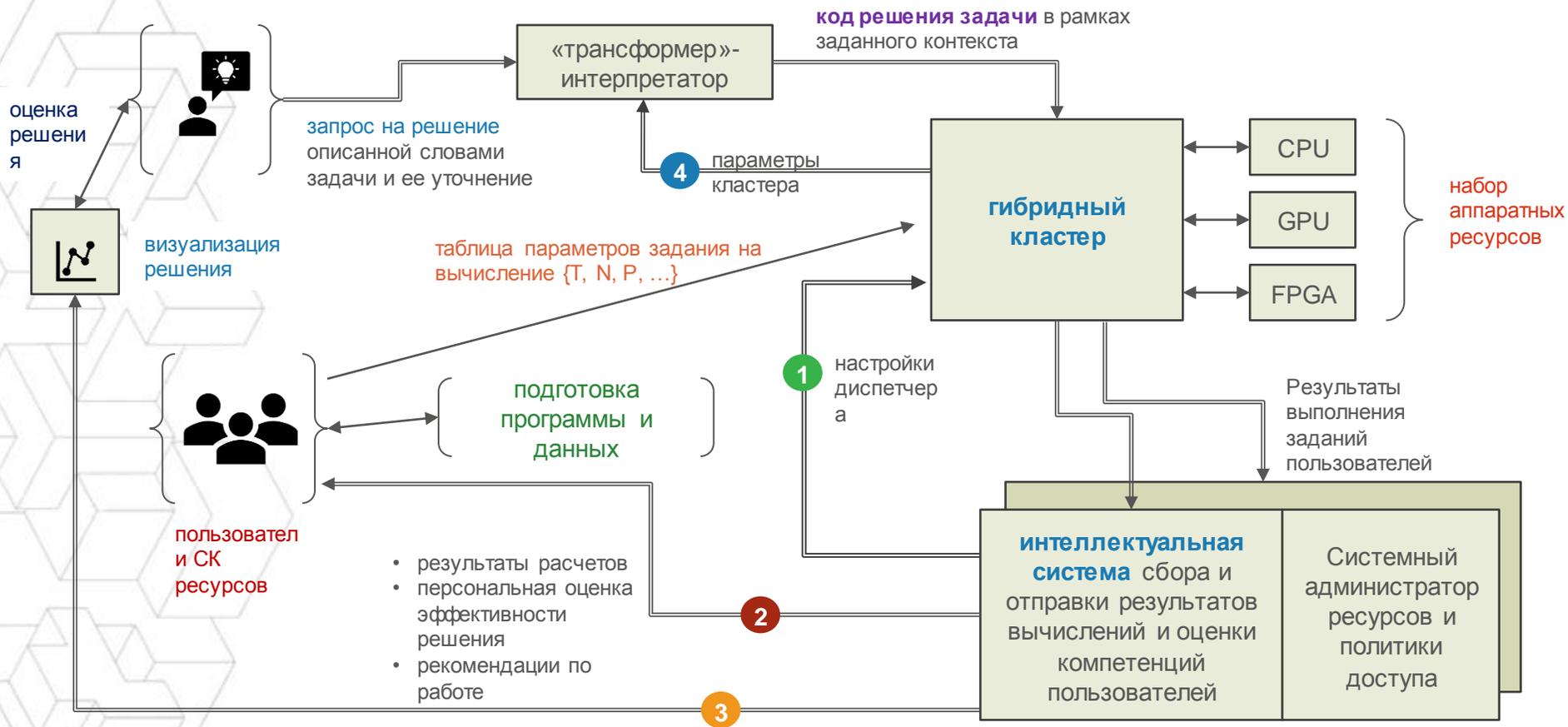
Konstantinov A. V., Utkin L. V. //arXiv preprint arXiv:2302.06601.— 2023.

РЕЗУЛЬТАТЫ ИСПОЛЬЗОВАНИЯ «УМНЫХ ОЦЕНОК» ВРЕМЕНИ РЕШЕНИЯ ЗАДАЧИ ПРИ РАСПРЕДЕЛЕНИИ РЕСУРСОВ ДИСПЕТЧЕРОМ SLURM





Что планируется сделать : РЕКУРСИВНАЯ КОГНИТИВНО-ДЕЯТЕЛЬНОСТНАЯ МОДЕЛЬ ПРОЦЕССОВ ВЫЧИСЛЕНИЙ В ЦКП «ПОЛИТЕХНИЧЕСКИЙ»



В пространстве «первых принципов» всегда много «свободного места»

Познание через эксперименты и опытные данные могут быть дополнены технологиями машинного обучения, решения которых валидируются «первыми принципами»

Перспективы использования моделей машинного обучения:

- позволяют эффективно использовать механизм внимания, веса которого определяют нормированное ядро непараметрической модели задания пользователя
- Ядро модели обучения может быть заменено на параметрическую функцию, значения параметров которой оптимизируются путём решения совместной задачи выживаемости и регрессии
- Для параметризации весов внимания могут использоваться различные подходы, в частности, многослойные нейронные сети



МОДЕЛИ "ВЫЖИВАЕМОСТИ" ПРИКЛАДНОЙ ЗАДАЧИ В СКЦ

Дано: Рассматривается «событие» - вероятность «успешного решения» задач пользователя до истечения заданного диспетчером времени T_i , которая связано с вектором x .

Имеется обучающая выборка (x_i, δ_i, T_i) , где вектор $x_i = (x_{i1}, \dots, x_{id})$, $i = 1, \dots, n$ - цензурированные данные о функционировании СК во время выполнения задания, включая параметры прикладной задачи, которые задаются самим пользователем

Заданы функции

- риска $h(t|x)$, что задача завершится в момент времени t
- кумулятивная функция риска $H(t|x)$,
- функция выживаемости $S(t|x) = \Pr\{T > t|x\}$
- (пример: модель пропорционального риска Кокса:
$$H(t|x, b) = H_0(t) \exp(xb^T)$$



Рассматривается модель выживаемости на основе случайного леса и обучающая выборка $(x_i, \delta_i, T_i), i = 1, \dots, n$.

Выход модели либо функция

- риска
- выживаемости (имеет не одно значение)

Требуется определить какие компоненты вектора x определяют функцию риска $H(t|x)$?

Например, функция риска на основе модели Кокса имеет вид:

$$H_{Cox}(t|x, b) = H_0(t) \exp(xb^T)$$



МОДЕЛИ ВЫЖИВАЕМОСТИ, УЧИТЫВАЮЩИЕ СОВМЕСТНОЕ ВЛИЯНИЕ ПРИЗНАКОВ

Имеется основная обученная модель выживаемости (случайный лес выживаемости, нейронная сеть) и датасет $(x_i, \delta_i, T_i), i = 1, \dots, n$.

Вопрос: какие пары признаков некоторого вектора x совместно определяют его функцию риска $H(t|x)$?

Идея: использовать метод Лассо и произведение признаков в расширенной модели Кокса:

$$H_{C_{0x}}(t|x, z, b, c) = H_0(t) \exp(xb^T + zc^T)$$

где z - вектор произведений пар признаков

$$z = (x_1x_2, x_1x_3, \dots, x_{d-1}x_d);$$

c - искомый вектор коэффициентов длины $d(d-1)/2$:

$$c = (c_{1,1}, c_{1,2}, \dots, c_{d-1,d})$$