Технологии машинного обучения в задаче управления ресурсами СК

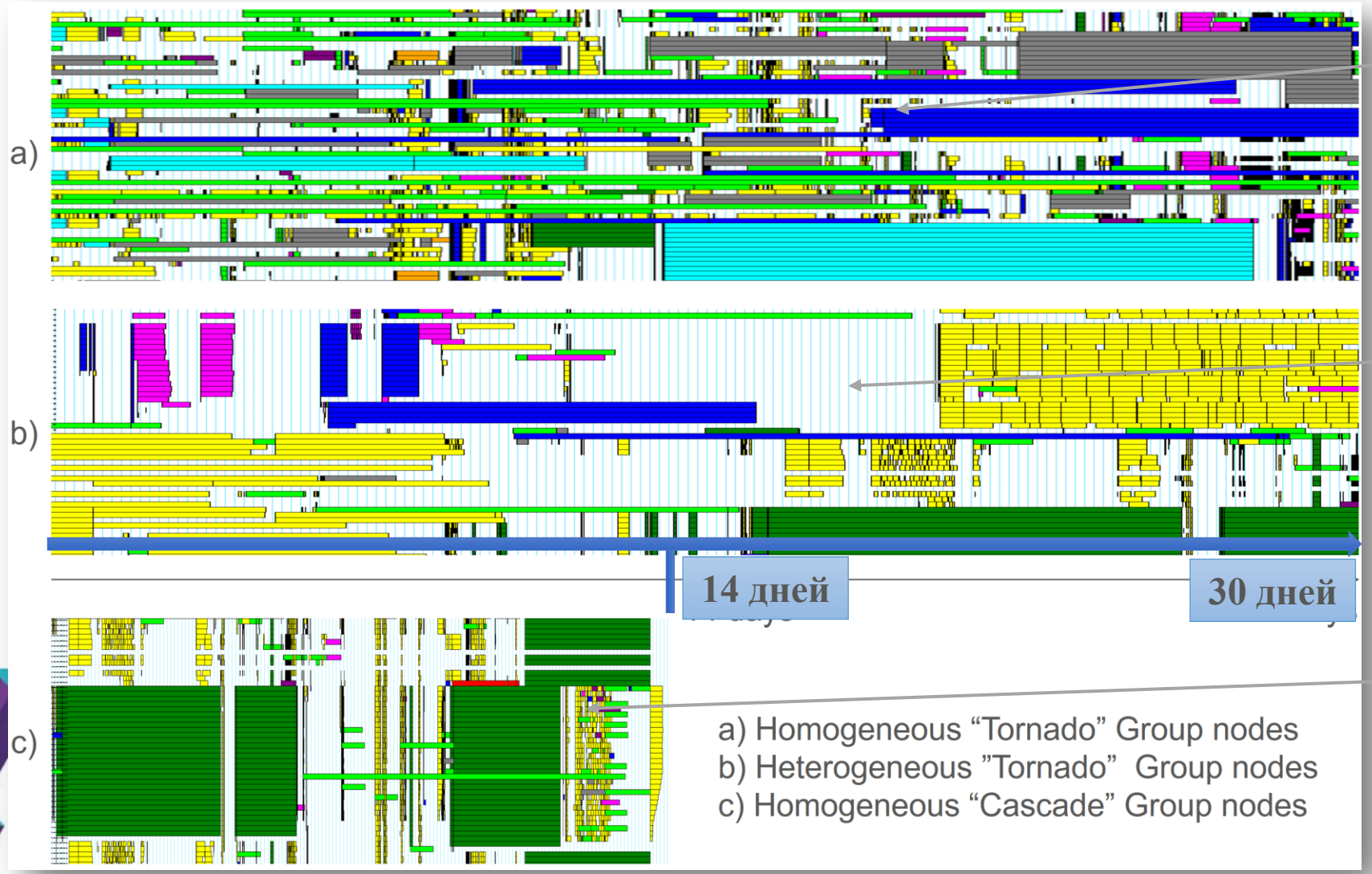
.

СКЦ «Политехнический»

14 февраля
2024 г.

ПОЛИТЕХ
Санкт-Петербургский политехнический университет Петра Великого

«Энтропия» процессов загрузки гибридного СКЦ «Политехнический»



Не более 60%

Не более 40%

Не более 50%

Потребляемые узлы, шт.

ПОЛИТЕХ

МЫСЛИТЬ БУДУЩИМ

Расщепление мира на себя и не-себя

Мир всегда, с первой секунды жизни, поделен на две части,
одна из которых направлена внутрь, а другая – наружу: ‘Я и все остальное’.
‘Мы и они. Свои и чужие.’ *)/// проблема исследований «СК и все остальное»

Осознание «себя» индивидуальностью – это функция СОЗНАНИЯ.

"*Сознание – это способность осознавать, оценивать себя отдельно от других*

Может ли осознать себя СК с ИИ - искусственный интеллектуальный агент?

"У животных есть осознание себя. Она не берется выполнять то, что не сможет сделать (например, перепрыгнуть канаву шириной в 3 метра)” **)

Может ли СК не исполнять «неправильно сформулированные задания» , а уточнять их параметры, чтобы «сделать их исполняемыми»

Объектом исследований являются - причинность и наблюдение:

- …«классы прикладных задач» и «компетенции пользователей», которые пытаются использовать конкретные вычислительные ресурсы в СК для своих целей.
- Предметом исследований выступают три рода сущностей : 1) время исполнения задачи в СК; 2) размерность …пространства факторов, влияющих на загрузку вычислительных ресурсов как со стороны задачи, так и со стороны порльзователей; 3) «к.п.д.» СК – как отношение числа успешно завершенных прикладных задач к общему числу загруженных в СК задач пользователей.
- Исследуются методы, посредством которых можно не только **описать** особенности прикладных задач и «поведение» пользователей, но также то, как используя полученные описания, перевести их в **«действия» (т.н. реализовать** процесс реификации):
  - то есть "продукты умственной деятельности» (концепции, модели, программы, представления) овеществить в той или иной форме.

  Пример: идея овеществление есть одна из часто используемых техник на тему "виртуализации" , например – мыслимы адрес, куда записывается драйвер устройства, превращается в «число»

  в C++ это реализуется так:

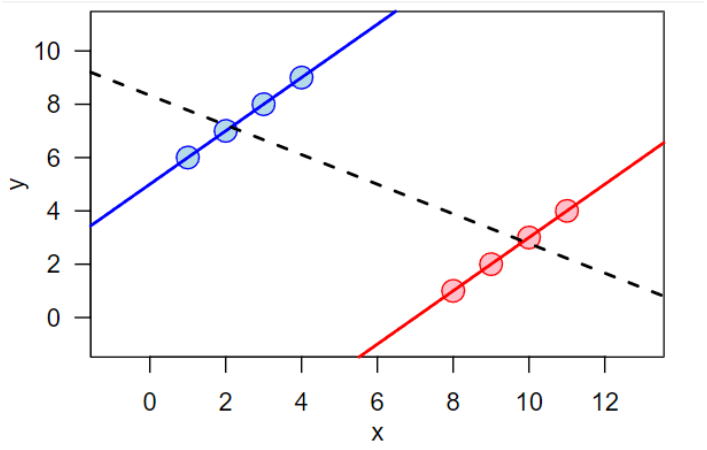
  «овеществление» физического адреса памяти для его непосредственного использования в каком либо специальном контроллере, чтобы указать место в памяти, куда будет записан драйвер:
  - char* buffer = (char*) 0xB8000000;
  - buffer[0] = 10;

характеристики приложений СКЦ как центра коллективного пользования

Проблема интерпретации многофакторных статистических моделей

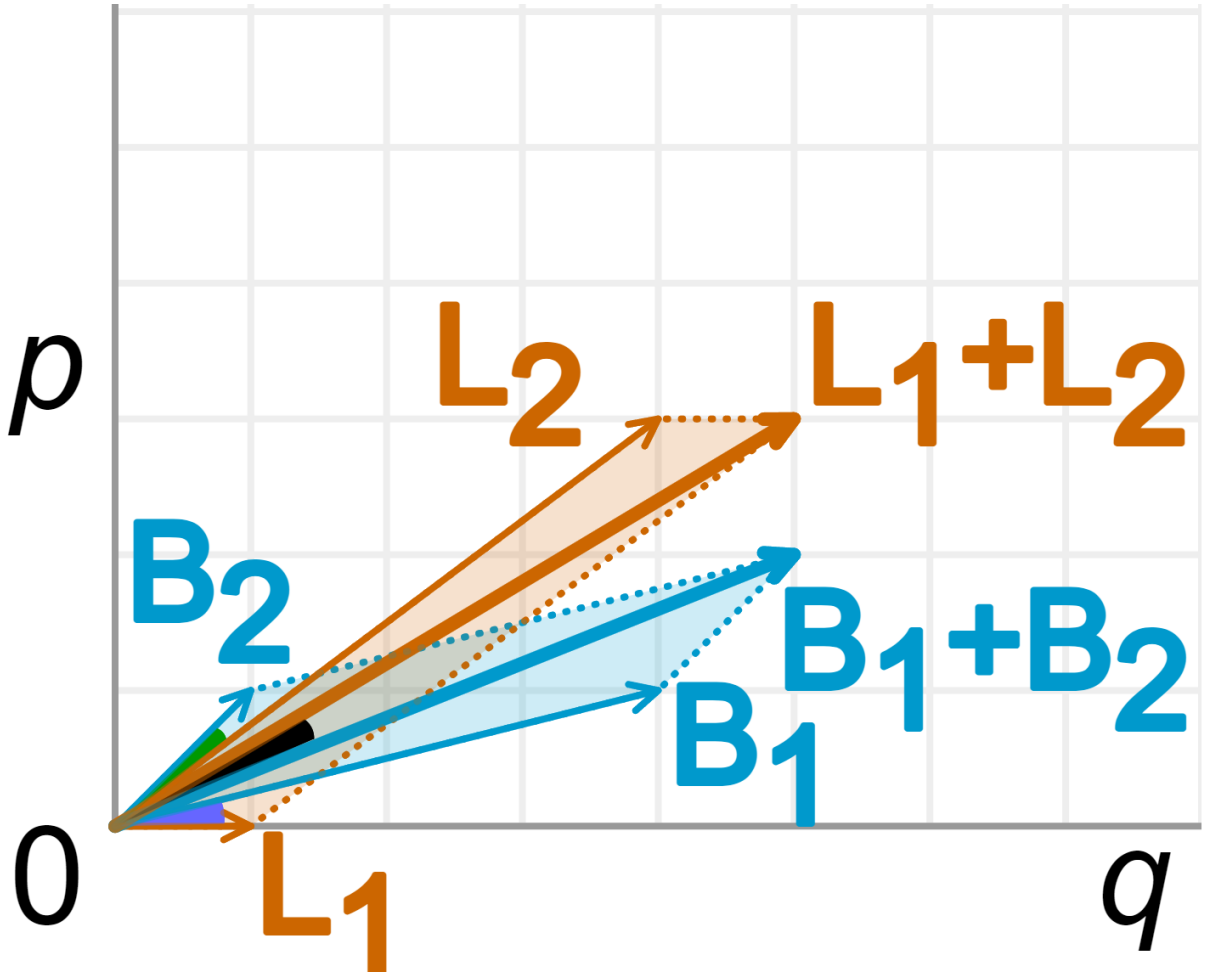
к статистическим данным следует относиться очень аккуратно

- Экспериментальное исследование «парадокса Симпсона» — контринтуитивное явление в статистике, когда мы видим в каждой из групп данных определенную зависимость, но при объединении этих групп зависимость исчезает или становится противоположной.
- Суть один и то же фактор по разному влияет на рассматриваемые группы факторов. Когда приводятся проценты реализации событий для нескольких групп данных, каждая из которых в свою очередь разбита на подгруппы, то количественная схожесть получившихся зависимостей (растет, уменьшается) требует дополнительного «причинно-следственного» объяснения



есть две положительные тенденции для разных подгрупп. Однако, если мы объединим данные, то получим тенденцию уже негативную (пунктирная линия).

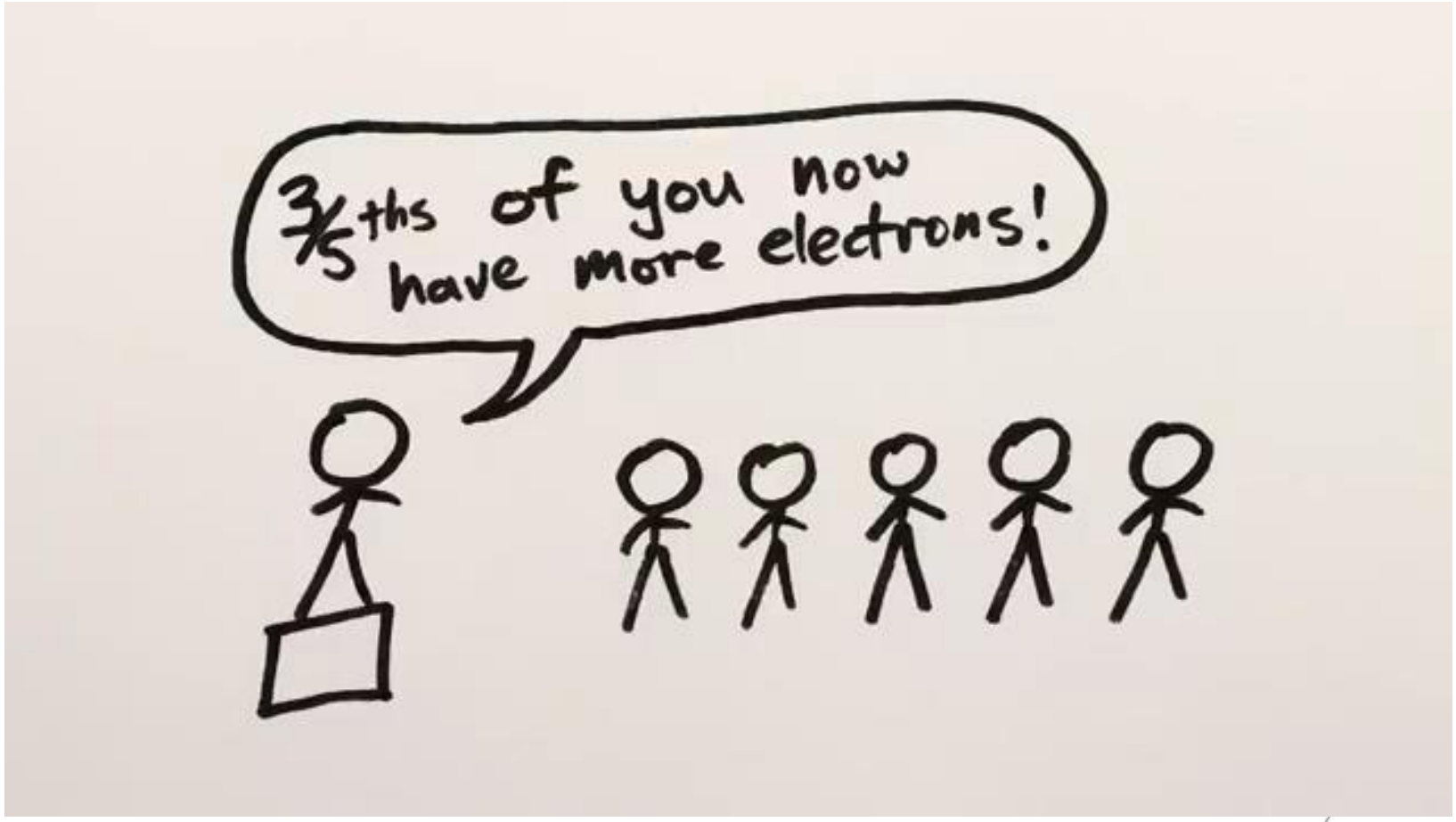
Объяснение «парадокса» - после - не значит вследствие.



На рисунке синим цветом представлены зависимости из одной подгруппы, оранжевой - из другой. B2 растет быстрее, чем L2 (наклон меньше), B1 растет быстрее, чем L1, однако векторная сумма говорит о том, что L1+L2 растёт быстрее!

Вывод:
Нужно находить именно причинно-следственные связи и на их основе выделять фактор-группы, для которых статистика будет на самом деле отражать текущие тренды.

Парадокс Симпсона



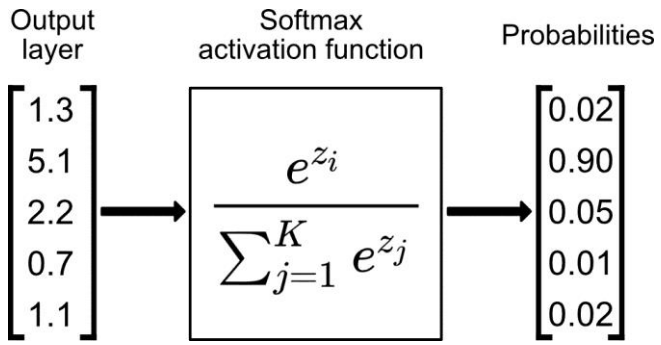
Организация исследований

- Построение сбалансированной выборки в каждом из выбранных классов
- Исследование Парадокс Симпсона —возникает, когда тенденция появляется в небольшом наборе данных, но эта тенденция исчезает или меняется на противоположную, когда набор данных делится на подгруппы.
- Исследование Парадокс Берксона - статистическое явление, при котором между между двумя переменными имеется отрицательная корреляция, а при разбиении данных на подгруппы или при отсутствии фактической корреляции между ними появляется положительная корреляция. Суть - парадокс возникает, когда независимые переменные имеют общую причину, которая не учитывается в модели.

  3 направления
  - 1 группа соревнуется между собой, используя и комбинируя известные модели
  - 2 группа разрабатывает критерии для тестирования и сравнения результатов первой группы
  - 3 группа свободные художники, осваивает объяснительные методы машинного обучения

Уточнения используемого понятия: «Механизм внимания»

В общем случае для описания сложной промышленной системы можно использовать непараметрические модели: $y = \sum_{i=1}^{N} \alpha(q, k_i) v_i$



в которых «ядро» модели

q – query - запрос

k_i – key - ключ

v_i – value – значение

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

В этом случае «механизмом внимания» называется преобразование действительных значений «ядра» модели $\alpha(q,k_i)$ в возможность (можно интерпретировать как вероятность прогнозируемых выходных классов) совершения некоторого действия.

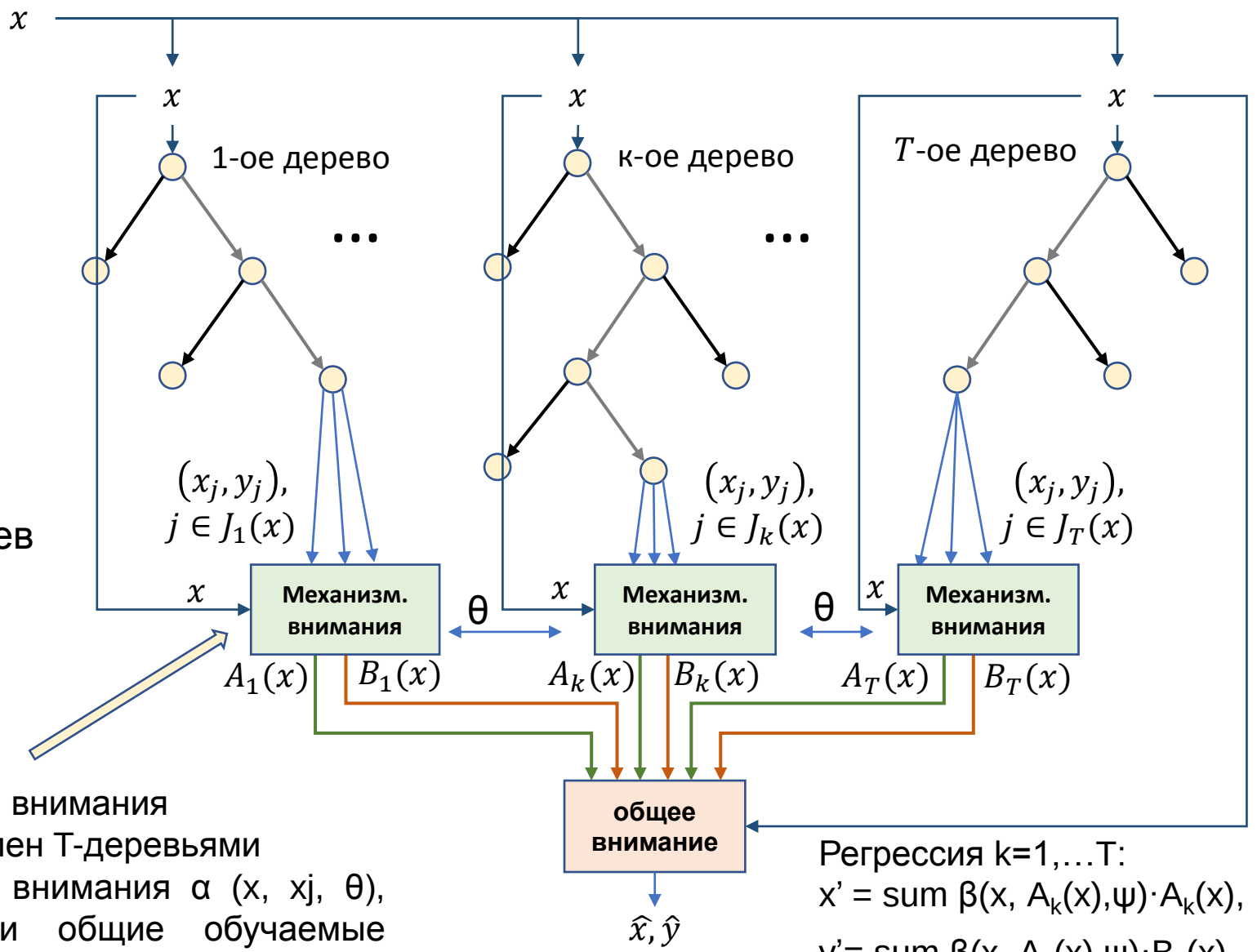
Комментарий: непараметрическое преобразование эффективны для

- задач моделирования сложных последовательностей (предложений естественного языка
- обработки произвольного числа пар сопряженных данных (ключ, значение)
- оптимизации функции $score(q, k_i) \propto q^T k_i$,

  либо $(W_Q q)^T (W_K k_i) = q^T W_Q^T W_K k_i = q^T W k_i$

Методы решения : использование «случайного леса» деревьев решений с механизмом внимания

Значительное повышение «объяснимой» точности решения **задачи классификации** достигается при использовании» гибридного» ансамбля («лес» +регрессия) деревьев решений с механизмом внимания («голосования»)



Механизм внимания представлен Т-деревьями с весами внимания α (x, xj, θ), имеющими общие обучаемые параметры θ

Регрессия k=1,…T:

x’ = sum $\beta(x, A_k(x), \psi) \cdot A_k(x)$,

y’= sum $\beta(x, A_k(x), \psi) \cdot B_k(x)$,

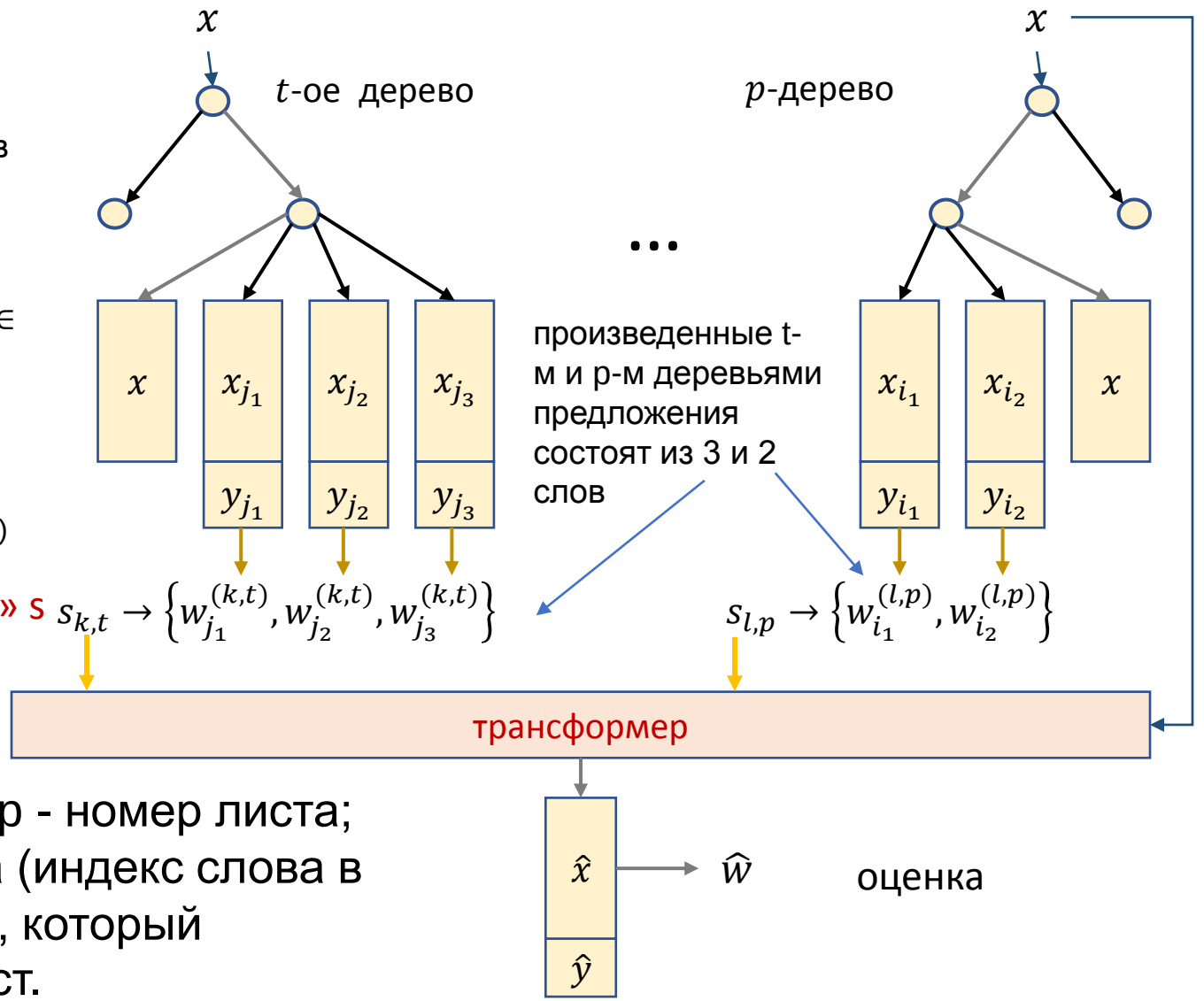
Новые методы решения : использование моделей трансформаторов совместно со «случайным лесом»

набор примеров (x_j, y_j), где $j \in J_k(x)$, которые попадают совместно с вектором признаков x в один и тот же p-й лист k-го дерева, представим в виде предложения s , состоящего из слов, обозначаемых как $w_i^{(k,p)} \in J_k(x)$,

$i \in J_k(x)$,

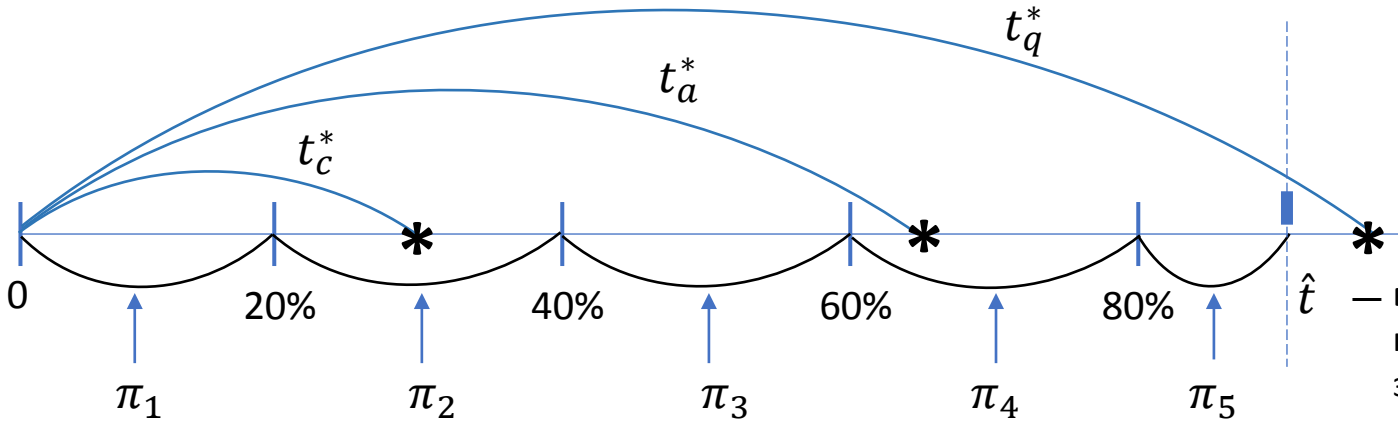
$s_{k,p} = \{w_{i1}^{(k,p)}, \ldots, w_{ip}^{(k,p)}\}$ $_{1\,ir}$

«Предложение» s



k - номер дерева; p - номер листа;
i - индекс примера (индекс слова в p-м предложении), который попадает в p-й лист.

Формирование обучающей выборки для оценки функции «выживания» прикладной задачи и функции «полезности диспетчера»



Почему уместен термин «выживание»: задача может завершиться «успешно» или быть «снята» диспетчером

$\hat{t}$ — время, заданное пользователем при постановке прикладной задания в очередь на исполнение

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

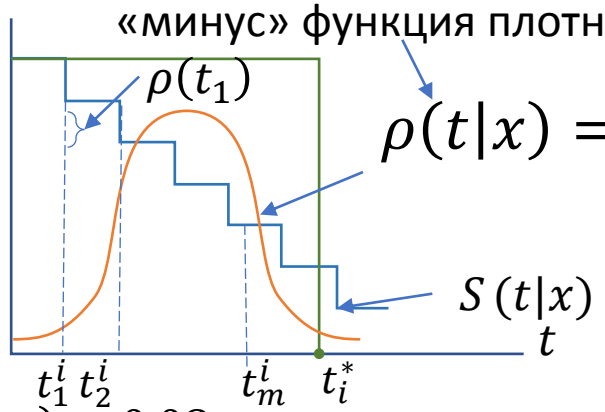
γ_1 задача q превысила время $\hat{t_q}$

γ_2 произошло прерывание (деление на 0)

… γ_m зацикливание задачи

$S(t|x)$

график функция выживаемости – аналог гистограммы распределения:



дискретные «приращения» графика в точках «событие»: $\rho(t_1) = 0.05$ $\rho(t_2) = 0.08$

Функция «полезности диспетчера»:

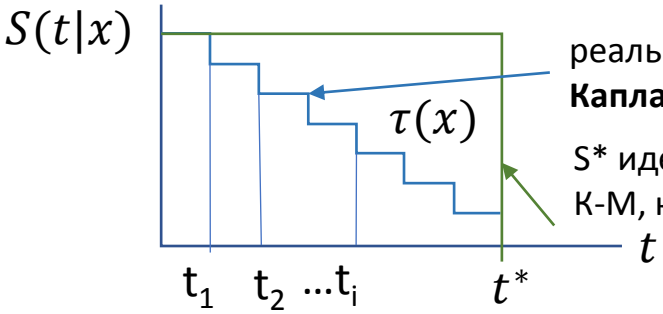
$$IU = \int_0^{t_{max}} u(t) \cdot \rho(t|x) dt$$

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

«абстракции», используемые для повышения реальной производительности СК :

Абстракции
а) «выживаемости» задач в СК
б) «полезности» диспетчера» СК

1. «Персональный» эффект от изменения функции выживаемости задания зависит от точности оценки времени решения t^*прикладной задачи



реальное значение **функции Каплана-Мейера**

S* идеальное значение функции К-М, когда прогноз пользователя совпадает с реальным временем задания выполнения

$S(t|x)$ «персональный» расчет для конкретного пользователя СКЦ

2. «Функция полезности» диспетчера СК характеризует «средний» эффект от загрузки узлов СК при условии **успешного завершения** «прикладного задания»

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

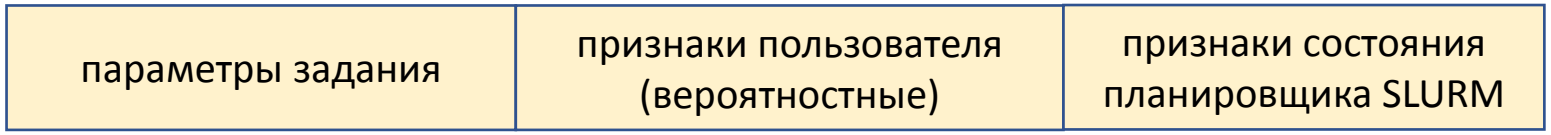
в случае диспетчера СК
$$IU = \sum_{i-\text{пользователь}}^{N} \sum_{j-\text{задача}}^{M} \int_0^{t_{max}(i,j)} u_{ij}(t)\rho_{ij}\left(t\middle|x_{ij}\right)dt$$

Структура аргументов функции «полезности» диспетчера СК

Функция полезности

$$\int_0^{t_{max}} u(t)\rho(t|x)dt \rightarrow \sum_{i-\text{пользователь}}^{N} \sum_{j-\text{задача}}^{M} \int_0^{t_{max}(i,j)} u_{ij}(t) \cdot \rho_{ij}\left(t\middle|x_{ij}\right)dt$$

Аргумент функции Выживаемости вектор x:



$$(\pi_{1i}, \dots, \pi_{ni}, \gamma_{1i}, \dots, \gamma_{1m})$$

Комментарий: часть вектор x при формирования заявки самому пользователю не известно. Эти данные необходимо оценивать в процессе работы СК.

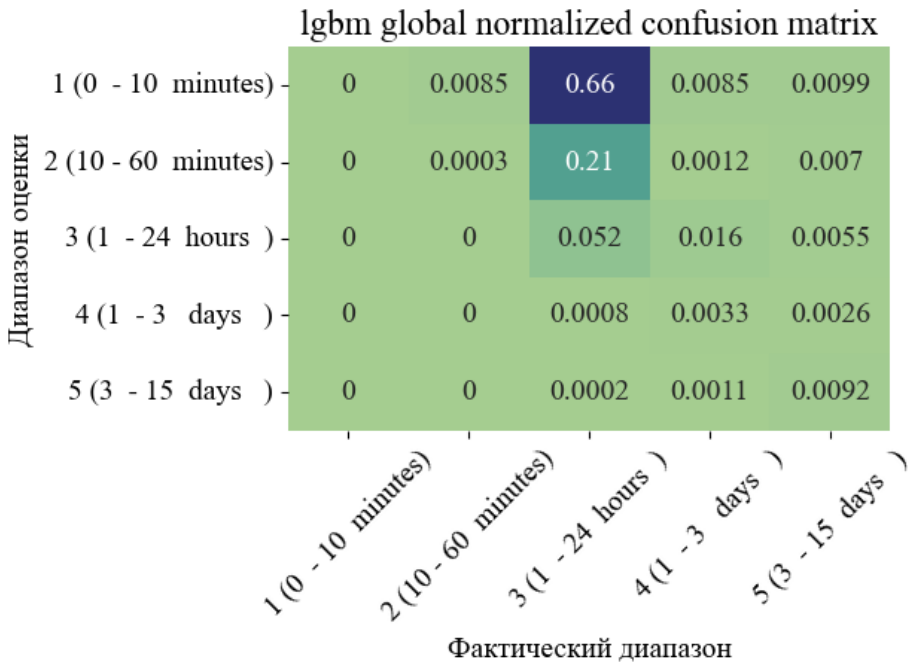
Вопрос: как это сделать, решалась ли эта задача раньше и в чем формальная сложность этой задачи/

Итак,

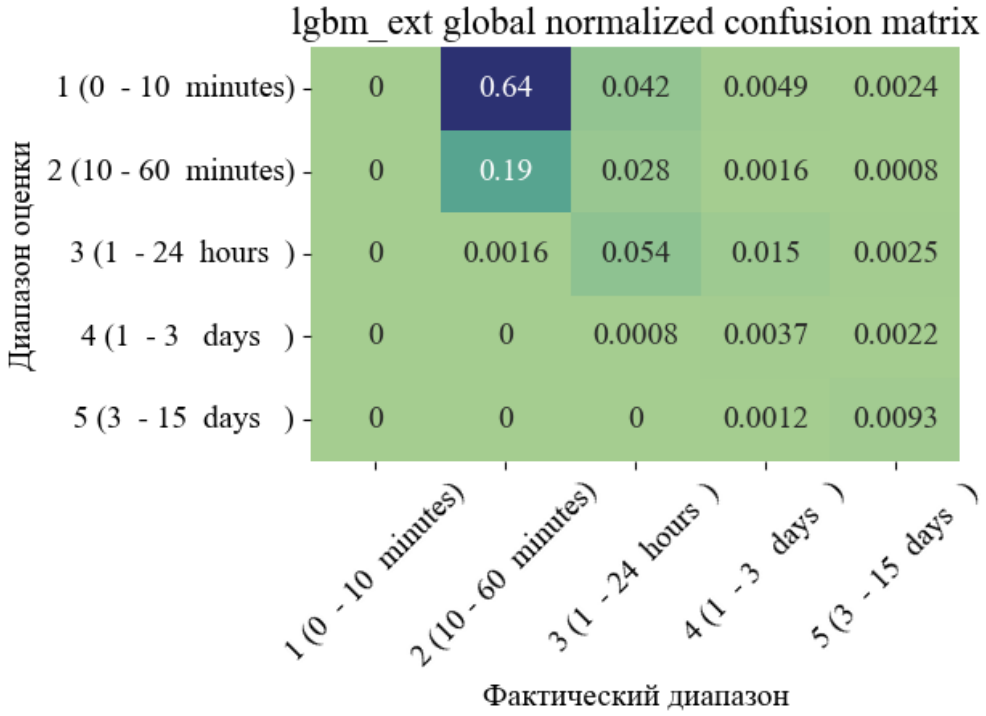
- Требуется построение моделей прогнозирования времени исполнения заданий в среде суперкомпьютерных вычислений
  - путем комбинации (гибридизация) различных моделей, таких как случайные леса, градиентный бустинг, трансформеры и нейронные сети…..

- Ps Построение «малых» моделей трансформеров (роя трансформеров) с механизмами внимания –**задача, которая может быть востребована** в современных СК системах, **использующих технологии ИИ.**

Формирование матриц плотностей распределения прогнозов обучаемой модели при оценке времени выполнения прикладной задачи.



Регрессия по информации о задаче



Регрессия по модели пользователя

Сравнение энтропийных оценок различных моделей прогноза (на глобальных матрицах)

Модель	r	r2	I = 1 - S	Eig values
Идеальная модель	1.00	1.00	0.86	[0.688 0.221 0.074 0.007 0.011]
Пользователь	0.46	-11.04	0.14	[0.000 0.000 0.020 0.002 0.011]
Модель среднего	0.00	0.00	0.19	[0.000 0.000 0.000 0.000 0.074]
Регрессия по информации о задаче	0.58	-0.12	0.22	[0.000 0.000 0.052 0.003 0.01]
Регрессия по модели пользователя	0.80	0.52	0.52	[0.000 0.191 0.054 0.003 0.01]