УДК 004.85

Раздельная оценка возрастного взвешивания обратной связи в предикторе и контроллере качества при маршрутизации языковых моделей

Исаев Ш. М.

Национальный исследовательский университет ИТМО

г. Санкт-Петербург, Российская Федерация

Маршрутизация запросов между языковыми моделями позволяет управлять соотношением стоимости и качества ответов [1, 2]. В потоковом сервисе оценка ответа поступает с задержкой [3]. В предшествующей работе автора (неопубликованные материалы; тезисы подготовлены к 69-й конференции МФТИ) исследован маршрутизатор DA-CB: оценка $q \in [0;1]$ выбранной модели приходит через d запросов с весом $w = e^{-a/\tau}\kappa^{[r_0 \neq r]}$, где a — возраст оценки, τ — масштаб, κ — множитель для оценок из другого обнаруженного режима. Вес применяется одновременно к ridge-обновлению предикторов качества и стоимости и к шагу двойственного множителя λ ограничения на потери. Там отключение возрастного множителя уменьшило среднюю потерю по всем 800 запросам в среднем по 27 условиям, включая $d = 0$ с нулевым эффектом («без возраста − полная версия»: −0,39 [−0,51; −0,27] п.п.); с оценками ниже эта величина не сравнима. Цель работы — выяснить, какой из двух механизмов связан с изменением потерь.

Метод. Контроллер: $\lambda \leftarrow \min\{\max[0,\ \lambda + \eta w_c((1-q)-\alpha)/\rho],\ \lambda_{\max}\}$, где α — допустимая потеря, ρ — доля запросов с оценкой; предиктор обновляется с весом w_p. Факторный план 2×2: $w_p \in \{\kappa^{[\cdot]},\ e^{-a/\tau}\kappa^{[\cdot]}\}$ (p0/p1), аналогично w_c (c0/c1); p1c1 — полная DA-CB, p0c0 — абляция без возраста. Контроль: $w_c = \kappa^{[\cdot]}$ с постоянным шагом $\eta_{eq} = \eta\, \mathrm{E}\, e^{-D/\tau}$. Остальные элементы и гиперпараметры ($\tau = 100,\ \kappa = 0{,}1,\ \eta = 0{,}05$) не менялись. Проверка

реализации: p1c1 и p0c0 воспроизвели решения исходных версий в 180 из 180 пар прогонов; при фиксированном d контроль и p0c1 дали 0 расхождений решений (420 пар), т. е. «возрастной» вес контроллера здесь тождествен уменьшению шага до $\eta e^{-d/\tau}$.

Данные и протокол. Офлайн-воспроизведение опубликованных оценок RouterBench [1]: MMLU, GSM8K, MBPP, HellaSwag, 11 моделей, исторические стоимости. В каждом из трёх сценариев за 400 запросами исходных семейств следуют 400 запросов целевого семейства (GSM8K, MBPP или HellaSwag), исключённого из обучения. Протокол зафиксирован до запуска; seeds 100–119 новые. Основное условие: $\rho = 1$, $d = 100$; вторичные: $\rho \in \{1;\ 0{,}25;\ 0{,}1\}$, $d \in \{20;\ 100\}$ и $D \sim U\{0, \ldots, 200\}$; с проверочным условием $d = 0$ и опорной моделью всего 2880 прогонов, неудачных нет. Исход V_{post} — средняя потеря $100(1 - q)$ на 400 запросах после смены семейства, п.п. Первичная оценка $E1 = V(\text{p0c1}) - V(\text{p1c0})$; также главные эффекты M_P, M_C, взаимодействие I и $TOTAL = V(\text{p1c1}) - V(\text{p0c0})$. Среднее по сценариям внутри seed; 95 %-е ДИ — парный бутстрэп 20 seed-блоков. Правило: компонент «отвечает» за изменение, только если ДИ его главного эффекта и ДИ $E1$ не содержат 0 при согласованном знаке.

Результаты (табл. 1).

Таблица 1. Оценки при $\rho = 1$, $d = 100$ (95 %-е ДИ)

Оценка	V_{post}, п.п.	Изменение средней λ после сдвига
$E1$	+0,31 [−0,20; +0,81]	−1,10 [−1,17; −1,02]
M_P	+0,19 [−0,05; +0,47]	0,00 [−0,01; +0,02]
M_C	+0,51 [+0,17; +0,86]	−1,10 [−1,17; −1,02]
I	+0,34 [+0,04; +0,65]	0,00 [−0,02; +0,01]
$TOTAL$	+0,70 [+0,36; +1,07]	−1,10 [−1,17; −1,02]

Версия с возрастным весом теряет больше ($TOTAL > 0$), но дешевле на 21,7 [18,9; 24,7] %. M_C отделён от 0 и сопровождается снижением λ и стоимости; M_P от 0 не отделён и на λ не влияет. Однако ДИ первичного контраста $E1$ содержит 0, а взаимодействие I положительно: вес предиктора повышает потери в основном вместе с весом контроллера. По заранее

заданному правилу механизмы не различены [3]. Для контекста: все варианты DA-CB дешевле опорной модели на 84,3–88,6 %, но их V_{post} равно 37,53–38,23 п.п. против 27,04 п.п.

Во вторичных условиях (эксплораторно, без поправки на множественность) условия правила для контроллера выполнены при $d = 100$, $\rho \in \{0{,}25;\ 0{,}1\}$ и при $\rho = 0{,}1$, $d = 20$ (M_C от +0,49 до +1,30 п.п.), но не в остальных; M_P нигде не отделён от 0.

Выводы. Предрегистрированный контраст не различил вклад предиктора и контроллера. Описательно возрастной вес контроллера снижает λ после сдвига и удешевляет маршрутизацию; рост потерь полной версии проявляется в основном при сочетании весов контроллера и предиктора ($I > 0$), а в первые 200 запросов после сдвига (вторичный исход) $E1 = +0{,}73$ [+0,08; +1,38] п.п. При постоянной задержке этот вес тождествен уменьшению шага, поэтому его корректнее рассматривать как выбор шага, а не как адаптацию к возрасту оценок; при переменной задержке отличие от постоянного шага по потерям не обнаружено (+0,06 [−0,16; +0,34] п.п.), хотя стоимость у повыборочного веса выше на 5,6 [4,2; 7,1] %. Это не доказательство механизма. Ограничения: офлайн-воспроизведение опубликованных оценок и исторических цен, дробные оценки GSM8K, одно фиксированное разбиение (ДИ условны), повторное использование пула целевых запросов исходной работы (анализ эксплораторный), ненастроенные гиперпараметры.

Список использованной литературы

1. Hu Q. J. *RouterBench: A Benchmark for Multi-LLM Routing System* / Q. J. Hu, J. Bieker, X. Li [et al.] // arXiv. – 2024. – arXiv:2403.12031. – URL: https://arxiv.org/abs/2403.12031 (дата обращения: 27.09.2026).
2. Ong I. *RouteLLM: Learning to Route LLMs with Preference Data* / I. Ong, A. Almahairi, V. Wu [et al.] // arXiv. – 2024. – arXiv:2406.18665. – URL: https://arxiv.org/abs/2406.18665 (дата обращения: 27.09.2026).
3. Howson B. Delayed Feedback in Generalised Linear Bandits Revisited / B. Howson, C. Pike-Burke, S. Filippi // PMLR. – 2023. – Vol. 206. – P. 6095–6119. – URL: https://proceedings.mlr.press/v206/howson23b.html (дата обращения: 28.09.2026).