

УДК 004.8

ОНЛАЙН-ОЦІНЮВАННЯ НАДІЙНОСТІ ДЖЕРЕЛ У ПОТОВОМУ АНАЛІЗІ МУЛЬТИМОДАЛЬНИХ ЧАСОВИХ РЯДІВ ІЗ КАЛІБРУВАННЯМ ІЗОТОНІЧНОЮ РЕГРЕСІЄЮ

І. С. Узун, М.В. Лобачев*Department of Artificial Intelligence and Data Analysis, Institute of Artificial Intelligence and Robotics, Odesa Polytechnic National University, Odesa, Ukraine*ORCID <https://orcid.org/0000-0001-6619-4862>ORCID <https://orcid.org/0000-0002-4859-304X>E-mail: uzun.i.s@op.edu.ua

АНОТАЦІЯ

Потокові інтелектуальні системи підтримки прийняття рішень, що обробляють мультимодальні часові ряди, працюють у режимі каузальності та мають задовольняти вимоги малої латентності, обмежених обчислювальних бюджетів і контрольованості реакцій на змінність середовища. Критичним практичним ризиком таких конвеєрів є тимчасова деградація окремих джерел (пропуски, підвищений шум, масштабні зсуви), яка може маскуватися під концептуальний дрейф та спричиняти нестабільні або надмірні керувальні дії. У роботі розглянуто онлайн-оцінювання надійності джерела як каузальної ймовірнісної оцінки перебування в недеградованому стані та показано, що для практичного керування потрібна саме калібрована шкала: значення на виході має інтерпретуватися як частота «недеградованого» режиму в релевантних умовах. Запропонований підхід поєднує легкі проксі-сигнали деградації, придатні для онлайн-обчислення, з калібруванням ізотонічною регресією, що забезпечує монотонне відображення шкору в коректну ймовірність. Ефективність оцінено метриками площі під ROC-кривою (ROC-AUC) для відокремлення деградованих станів та очікуваної помилки калібрування (Expected Calibration Error; ECE) для контролю ймовірнісної узгодженості. Ключові експериментальні результати демонструють ROC-AUC $0,86 \pm 0,07$ для каліброваного варіанта та покращення каліброваності від ECE = $0,18 \pm 0,07$ (без калібрування) до ECE = $0,08 \pm 0,04$ (після калібрування) за прийнятних часових витрат: прості шкали-проксі мають мікросекундні затримки, а повна онлайн-модель зберігає середню латентність на рівні близько $150 \mu s$, що відповідає потребам поточкових конвеєрів. Отриманий результат формує інтерпретований керувальний сигнал, придатний для інтеграції з процедурами злиття, подієвої бюджетованої адаптації та контрольованої сигналізації в поточкових системах.

Ключові слова: машинне навчання, аналіз даних, інформаційні системи, системи підтримки прийняття рішень, мультимодальні часові ряди, поточковий аналіз, онлайн-калібрування, ізотонічна регресія, виявлення деградації, оцінювання надійності.

Вступ

Потокові інтелектуальні системи підтримки прийняття рішень інтегрують дані з різномірних джерел і формують прогнози та сигнали стану в режимі, наближеному до реального часу. У багатьох прикладних доменах дані природно мають вигляд часових рядів, а різні сенсорні та інформаційні канали утворюють мультимодальний потік, у якому модальності відрізняються масштабами, частотами дискретизації, затримками та типовими дефектами вимірювання. Для такого класу систем вирішальними є каузальність прийняття рішень, контроль латентності та керуваність обчислювальних витрат, оскільки саме

ці фактори визначають можливість експлуатації алгоритмів у потоці.

Практичний збійний сценарій полягає в тимчасовій деградації окремих модальностей: зростанні частки пропусків, підсиленні шуму або появі масштабних зсувів, що порушують стабільність прогнозування та сигналізації. У потоці ці ефекти часто накладаються на нестационарність середовища (зокрема, концептуальний дрейф), тому зростання похибки або зміна скорів може бути неоднозначно інтерпретоване. У результаті система ризикує або «пропускати» реальні відмови джерел, або навпаки запускати надмірні коригувальні дії (адаптацію, перемикання

режимів, перезважування модальностей), що підвищує вартість експлуатації та знижує довіру до рішень.

Одним зі способів зробити реакції потокової системи підтримки прийняття рішень керованими є введення явного *керувального сигналу якості* модальності, який відокремлює деградацію джерела від інших причин нестабільності. Однак для практичного використання цей сигнал має бути не лише дискримінативним, а й *інтерпретованим* та *каліброваним*, інакше пороги та правила керування не матимуть стабільного смислу між різними сегментами потоку. Саме ця мотивація визначає фокус статті: онлайн-оцінювання надійності з калібруванням ізотонічної регресією за збереження мікросекундного порядку латентності.

Аналіз літературних даних і постановка проблеми

Мультимодальне навчання та злиття даних є предметом систематичних оглядів, у яких запропоновано таксономію мультимодальних задач (representation, translation, alignment, fusion, co-learning) і підкреслено, що різна якість та неповнота модальностей є типовими проблемами інтеграції [1]. Огляд з мультимодального data fusion узагальнює стратегії об'єднання та формулює ключові виклики (узгодження, невизначеність, неповнота, неоднорідність), через які робастна інтеграція потребує явного врахування характеристик кожного джерела [2].

Для поточних сценаріїв принциповим є коректне онлайн-оцінювання в каузальних протоколах. Передпослідовний (prequential) підхід формалізує узгоджений режим test-then-train для хронологічних даних і використовується як базова методологія оцінювання в поточних задачах [3]. Практичні керівництва з навчання на потоках даних (зокрема, з прикладами в MOA) описують типові алгоритми, протоколи валідації та інженерні обмеження реального потоку [4]. Нестационарність потоків у прикладних задачах часто пов'язують із concept drift, для якого запропоновано класифікації типів дрейфу (раптовий, поступовий, рекурентний) та підходи до адаптації [5, 6]; окремо наведено порівняльний аналіз ансамблевих стратегій, що демонструє залежність ефективності від характеру нестационарності [7].

Окремою лінією досліджень є оцінювання й інтерпретація невизначеності прогнозів. У байєсівському глибокому навчанні обґрунтовано розрізнення алеторної та епістемічної невизначеності й описано практичні схеми їх оцінювання [8], а також показано, що під зсувом даних оцінки невизначеності можуть ставати ненадійними [9]. Оглядові роботи пропонують уніфіковану формальну рамку для трактування різних типів невизначеності в машинному навчанні [10]. Проте невизначеність моделі та якості джерела

даних не є тотожними: тимчасова деградація сенсора є властивістю вхідних спостережень і повинна діагностуватися каузально в термінах стану джерела, щоб керування було аудитованим.

У цьому контексті центральним стає калібрування ймовірнісних шкал. Показано, що навіть точні моделі (зокрема, сучасні глибокі мережі) можуть бути систематично некаліброваними, а temperature scaling є простою практичною корекцією; для контролю каліброваності широко застосовують ECE та діаграми надійності [11]. Класичне сигмоїдне (логістичне) калібрування перетворює скорі класифікатора на ймовірності та використовується, зокрема, для побудови ймовірнісних виходів SVM [12]. Непараметричне калібрування ізотонічною регресією будує монотонне відображення «скор \rightarrow ймовірність» без параметричних припущень [13], а експериментальні порівняння показують, що цей підхід є гнучким, але чутливим до розміру калібрувального набору [14]. Бета-калібрування розглядається як теоретично обґрунтована альтернатива логістичному калібруванню для бінарних класифікаторів [15]. Окремо підкреслено, що оцінювання каліброваності потребує коректних статистичних процедур: стандартна ECE може бути зміщеною, а kernel-based тести пропонуються як більш строгий інструмент [16].

Для поточних систем важливо, щоб калібрування не повинно порушувати часові бюджети. Класичний ROC-аналіз обґрунтовує ROC-AUC як узагальнену міру якості ранжування, інваріантну до вибору порога, що є критичним для поточних сценаріїв із динамічними порогоми [17].

Проблема полягає у відсутності підходу, який одночасно забезпечує (i) дискримінативне відокремлення деградованих станів, (ii) калібровану ймовірнісну інтерпретацію шкали надійності та (iii) мікросекундний порядок латентності, придатний для поточних конвеєрів.

Мета та задачі дослідження

Метою дослідження є розробка й експериментальна перевірка підходу до онлайн-оцінювання надійності джерел у поточному аналізі мультимодальних часових рядів на основі легких проксі-сигналів деградації з калібруванням ізотонічною регресією. Для досягнення мети поставлено такі задачі:

- 1) формалізувати інтерпретацію надійності як керувального інтерфейсу потокової системи підтримки прийняття рішень;
- 2) обґрунтувати роль калібрування для стабільного використання порогів у правилах керування;
- 3) експериментально оцінити якість підходу за показниками ROC-AUC, ECE та латентності;
- 4) визначити обмеження й умови застосовності в реальних поточних системах.

Постановка задачі. Нехай $\{x_t^{(m)}\}_{t \in \mathbb{Z}}$ – мультимодальний потік, де $m \in \{1, \dots, M\}$ індексує модальність, а t – дискретний час. Деградація модальності розглядається як тимчасовий стан, що зумовлює погіршення інформативності або коректності спостережень, і може проявлятися у вигляді пропусків, шуму чи масштабних зсувів. Задача онлайн-оцінювання надійності полягає у формуванні в кожний момент часу t числа $r_t^{(m)} \in [0, 1]$, яке інтерпретується як імовірність перебування модальності в недеградованому стані за каузально доступною історією.

Ключовою прикладною вимогою є каліброваність цієї шкали. Тобто якщо система повідомляє $r_t^{(m)} \approx 0,8$, то в довготривалому горизонті в схожих умовах частка «недеградованого» стану має бути близькою до 80%; інакше пороги та політики керування стають неаудитованими. Якість відокремлення деградаційних станів оцінюється за ROC-AUC, а каліброваність – за ECE. Додатково для потокового режиму потрібна перевірка латентності як характеристика застосовності: метод має вкладатися в мікросекундні або субмілісекундні бюджети на крок, щоб не порушувати загальні вимоги до рівня сервісу конвеєра.

Легкі проксі-сигнали деградації. У потоковому режимі прямі «важкі» діагностики якості часто є недоступними або занадто дорогими, тому доцільно використовувати проксі-сигнали, що обчислюються каузально та мають низьку обчислювальну складність. У запропонованому підході акцент зроблено на проксі, пов'язані з часткою пропусків та зі статистиками, що відображають зміну похибки або штрафів, які супроводжують прогнозування. Ці сигнали не є повними моделями деградації, але вони придатні як оперативні індикатори в системах, де головною вимогою є швидкість реакції та контроль витрат.

Онлайн-оцінювання та калібрування. Базова ідея полягає в тому, що початковий скор надійності, побудований на проксі-сигналах, може бути дискримінативним, але некаліброваним. Щоб перетворити такий скор на ймовірність з інтерпретованим смислом, застосовується калібрування ізотонічною регресією, яка будує монотонне відображення «скор \rightarrow імовірність». Монотонність є принциповою: вона узгоджується з інтуїцією проксі-сигналів деградації (гірший сигнал не повинен підвищувати надійність) і зменшує ризик артефактів керування, коли пороги починають реагувати нестабільно.

Практична цінність калібрування проявляється в тому, що пороги та правила, налаштовані на каліброваній шкалі, зберігають смисл між різними сегментами потоку та допускають аудит. З огляду на це надійність можна трактувати як керувальний інтерфейс: вона може використовуватися для переважування модальностей у злитті, для умовного запуску подієвих процедур адаптації та

для стабілізації сигналізації в задачах контролю ризику.

Порівняння ROC-кривих для різних варіантів оцінювання надійності наведено на рис. 1.

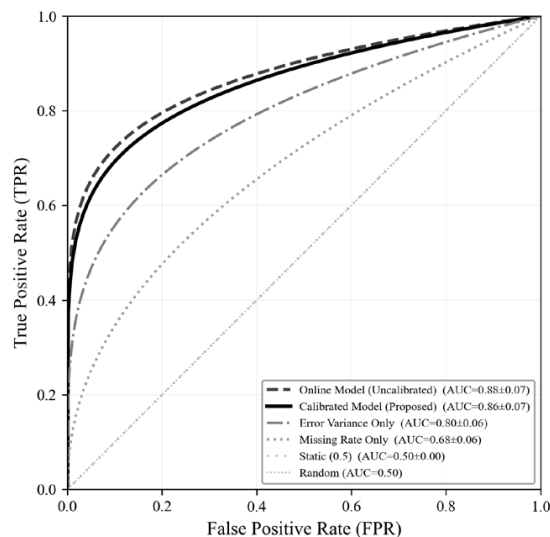


Рис. 1. ROC-криві для різних варіантів оцінювання надійності

Метрики якості та їх інтерпретація. Для дискримінаційної здатності шкали надійності у відокремленні «деградовано / недеградовано» доцільною є ROC-AUC як узагальнена міра якості ранжування, інваріантна до вибору конкретного порога [17]. Це особливо важливо в потокових сценаріях, де поріг може вибиратися на основі «чистого» префікса, під обмеження на частку хибних тривог, або як функція від бюджетів. Водночас ROC-AUC не відповідає на питання, чи можна інтерпретувати $r_t^{(m)}$ як імовірність; для цього потрібні калібрувальні метрики.

Метрика ECE вимірює узгодженість між «довірою» (confidence) та фактичною частотою правильних подій у групах спостережень. Один із поширених варіантів визначення ECE має вигляд:

$$ECE = \sum_{b=1}^B \frac{|I_b|}{n} |\text{acc}(I_b) - \text{conf}(I_b)|, \quad (1)$$

де n – кількість спостережень у калібрувальному наборі; B – кількість бінів; I_b – індекси спостережень, що потрапили в b -й бін за значенням скору; $\text{acc}(I_b)$ – емпірична частота «позитивної» події (наприклад, недеградованого стану) в біні; $\text{conf}(I_b)$ – середнє значення скору в біні [9]. Низька ECE означає, що шкалу можна використовувати як керувальну ймовірність у правилах, де важливе узгодження порогів із реальною частотою подій.

Діаграми надійності для каліброваного, некаліброваного й одноозначового варіантів наведено на рис. 2.

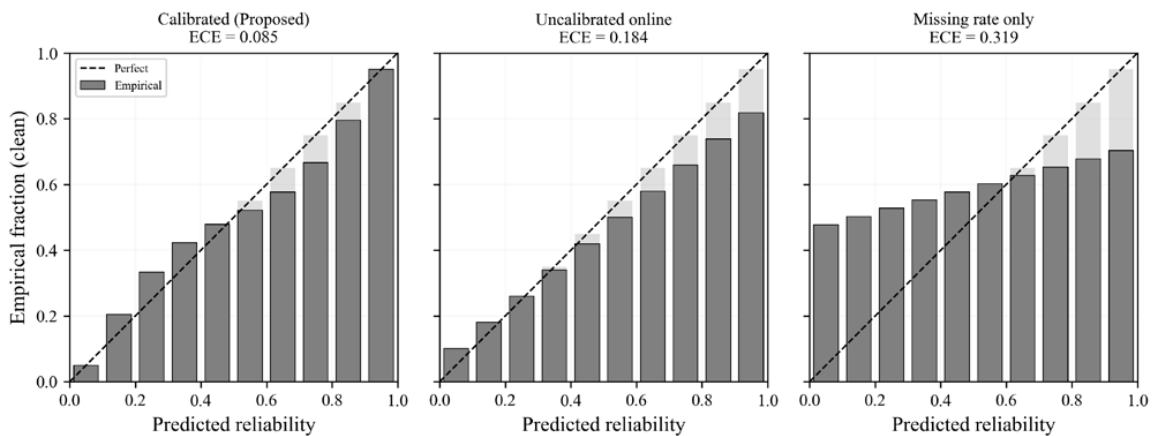


Рис. 2. Діаграми надійності (reliability diagrams) для каліброваного, некаліброваного й однознакового варіантів

Калібрування ізотонічною регресією. Нехай s_i – некалібрований скор (або логіт / скалярний скор) для i -го прикладу, а $z_i \in \{0,1\}$ – мітка «недеградовано / деградовано» для цього прикладу в калібрувальному вікні. Ізотонічна регресія будує монотонне відображення $g(\cdot)$, що мінімізує квадратичну помилку за умов монотонності:

$$g^* = \arg \min_{g \in \mathcal{G}} \sum_{i=1}^n (z_i - g(s_i))^2 \text{ за умови, що } g \in \text{неспадною}, \quad (2)$$

де \mathcal{G} – клас монотонних (неспадних) функцій [13; 14]. Побудоване g^* застосовується до скорів у потоці, формуючи калібровану оцінку надійності $r_t = g^*(s_t)$. З погляду керування це означає, що шкала r_t узгоджена з емпіричною частотою недеградованих станів у калібрувальних умовах, а монотонність гарантує відсутність «перевертання» порядку: більш «поганий» скор не може перетворитися на вищу ймовірність.

У потоковому застосуванні ізотонічне калібрування є привабливим тим, що (i) не потребує перевизначення базової моделі та може виконуватися поверх довільних скорів; (ii) не вимагає параметричного припущення щодо форми перетворення; (iii)

узгоджується з інтуїцією проксі-ознак деградації, де порівняльний порядок часто надійніший за абсолютну шкалу. Разом із тим, як і будь-яке калібрування, ізотонічна регресія залежить від репрезентативності калібрувального сегмента та потребує моніторингу під час експлуатації.

Загальну схему конвеєра онлайн-оцінювання надійності наведено на рис. 3.

Калібрована надійність як керувальний інтерфейс у потоковій системі підтримки прийняття рішень. Практична мотивація введення $r_t^{(m)}$ полягає не в самій по собі діагностиці, а в тому, що шкала повинна бути придатною для перетворення на дискретні дії, які мають чітку експлуатаційну інтерпретацію. У потокових конвеєрах типові керувальні дії можна умовно поділити на три групи. Перша група – *реакції на деградацію даних*: пригнічення або відключення проблемного каналу, перехід до більш надійної модальності, сигнал оператору про потребу перевірки сенсора. Друга група – *реакції на нестаціонарність середовища*: запуск обмеженої за бюджетом адаптації, зміна параметрів згладжування або порогів сигналізації, період «охолодження» після

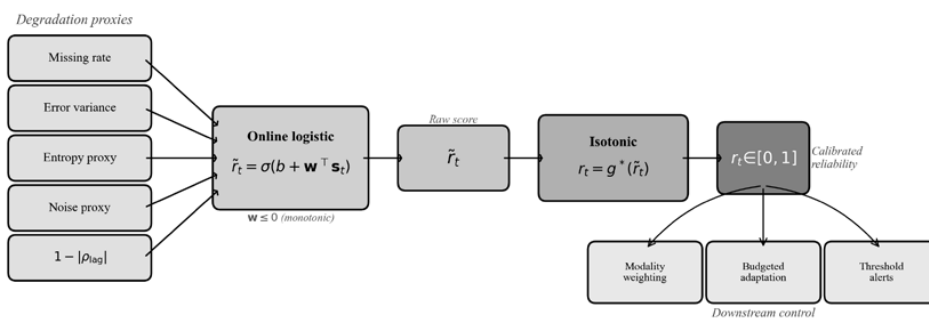


Рис. 3. Схема конвеєра онлайн-оцінювання надійності

адаптації. Третя група – *реакції на ризики аномалій*: керування режимом скорингу та порогами за фіксованого бюджету хибних тривог. У всіх цих випадках ключовою умовою є те, що пороги повинні мати стабільний смисл, і саме тому каліброваність шкали є системоутворювальною вимогою.

Калібрована ймовірнісна інтерпретація $r_i^{(m)}$ дає можливість задавати правила керування, прив'язані до зрозумілих інваріантів. Наприклад, поріг $r_i^{(m)} \geq 0,9$ може інтерпретуватися як «модальність вважається надійною з імовірністю не нижче за 0,9», а поріг $r_i^{(m)} < 0,5$ – як «переважно деградований режим», що допускає перехід до консервативної політики. Без калібрування схожі правила перетворюються на евристику, яка може працювати на одному запуску та руйнуватися на іншому через дрейф шкали або зміни розподілу проксі-ознак.

Режими доступності міток деградації та експлуатаційний моніторинг. Практична реалізація калібрування залежить від того, чи доступні мітки деградації z_i під час роботи системи. У контрольованих експериментах або стендових тестах мітки можуть бути сформовані детерміновано (через ін'єкції деградацій), що дає змогу калібрувати шкалу й оцінювати ЕСЕ без додаткових припущень. У реальних системах підтримки прийняття рішень мітки можуть бути частковими (наприклад, за сервісними подіями сенсорів, ручною верифікацією оператора або післядією технічних інцидентів) або взагалі недоступними. У цих умовах ключовим стає експлуатаційний моніторинг: відстеження стабільності розподілу r_i , контроль частоти переходів між режимами, аналіз кореляції з відомими інцидентами, а також періодична повторна калібровка на сегментах, де мітки можна отримати. Саме поєднання «калібрування + моніторинг» робить шкалу відтворюваним інтерфейсом керування, а не одноразово налаштованим скором.

Порівняння латентності (середньої та p99) для різних варіантів оцінювання надійності наведено на рис. 4.

Часові бюджети та мікробенчмарк. У потоковому режимі обчислювальні витрати модуля надійності повинні бути співмірні з витратами основного предиктора і не повинні створювати «вузьке місце» в конвеєрі. Тому оцінювання латентності розглядається як обов'язкова складова валідації: навіть помірне збільшення затримки на крок може акумулюватися на довгих горизонтах і порушувати вимоги до оновлення рішень. Наведені в табл. 2 значення демонструють характерну картину компромісу: прості проксі-шкали є значно швидшими, тоді як повна онлайн-модель забезпечує більш інформативну та калібровану шкалу за середньої латентності близько $150\mu s$. Для практики важливо, що обидва режими перебувають у мікросекундному діапазоні, а отже, можуть бути інтегровані в потокові конвеєри без радикальної перебудови архітектури.

Результати абляційного аналізу, що відображають внесок окремих проксі-ознак, наведено на рис. 5.

Прості проксі-шкали привабливі мінімальними витратами, однак їхня прикладна корисність часто обмежується вузьким класом деградацій. Наприклад, частка пропусків безпосередньо сигналізує про втрату даних, але може не реагувати на шумові деградації або масштабні зсуви; статистики похибки можуть бути чутливими до деградації, але потребують доступності фактичної цілі y_i та можуть бути конфундовані з дрейфом задачі. Тому практичний компроміс полягає в комбінуванні кількох легких сигналів і в перетворенні сумарного скору в калібровану шкалу. Калібрування в цьому контексті виконує роль «нормалізатора смислу»: воно робить пороги та правила керування відтворюваними і зменшує ризик, що та сама числова межа означатиме різні рівні ризику на різних сегментах потоку.

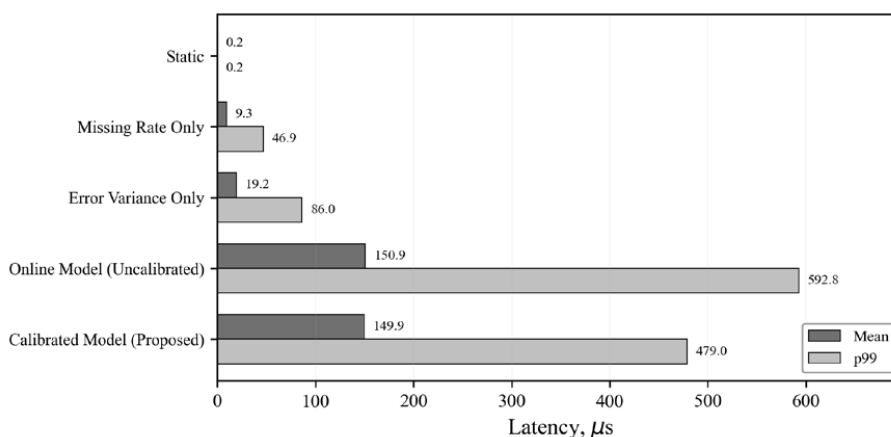


Рис. 4. Порівняння латентності (середня та p99) для різних варіантів оцінювання надійності

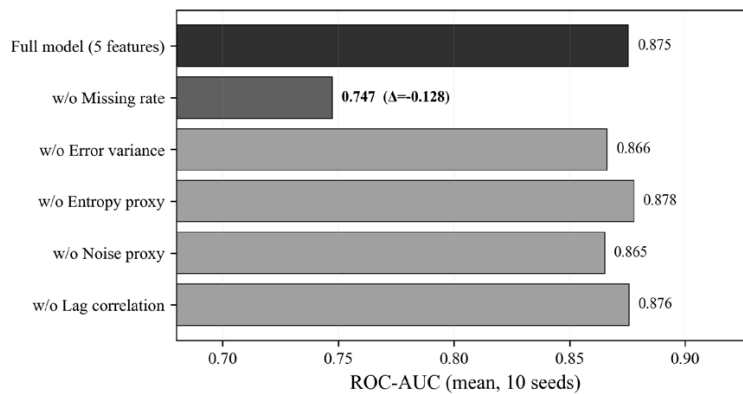


Рис. 5. Вплив окремих проксі-ознак на якість оцінювання надійності (абляційний аналіз)

Калібрувальний сегмент та оновлення відображення. Ізотонічна регресія, як і будь-який калібратор, потребує вибору калібрувального сегмента, у якому скор s_t та мітки z_t є репрезентативними щодо очікуваних експлуатаційних умов. У лабораторному сценарії таким сегментом може бути контрольований потік із детермінованими ін'єкціями деградацій; у виробничому – відрізки, підтверджені сервісними подіями сенсорів або ручною верифікацією. У випадку зміни домену або механізмів деградації потрібне повторне калібрування або розширення калібрувального набору. Практичний підхід полягає в тому, щоб розглядати калібрування як частину експлуатаційного циклу: збір нових прикладів, перевірка узгодженості (за ECE на доступних мічених сегментах) та оновлення g^* за потреби.

Рекомендації щодо впровадження в потоковий конвеєр. Для перенесення запропонованого підходу в прикладну систему підтримки прийняття рішень доцільно дотримуватися такого мінімального регламенту:

1) визначити прикладне поняття «деградованого стану» для кожного джерела та перелік типових деградацій (пропуски, шум, зсуви);

2) вибрати набір легких проксі-сигналів, які каузально обчислюються в потоці та відображають ці деградації;

3) сформувані калібрувальний сегмент із мітками деградації (контрольований тест, журнал інцидентів, ручна розмітка) і побудувати ізотонічне відображення;

4) визначити пороги керування на каліброваній шкалі r_t та правила моніторингу (частота перемиць, частка часу в «ненадійному» режимі);

5) періодично перевіряти узгодженість шкали на доступних мічених сегментах і оновлювати калібрування в разі зсуву умов.

Зазначений регламент не є складним з інженерного погляду, але суттєво підвищує аудиторність і стійкість рішень у потоці.

Експериментальні показники якості та латентності. Табл. 1 і табл. 2 узагальнюють ключові метрики якості та часові характеристики запропонованого підходу.

Табл. 1. Узагальнення якості оцінювання надійності за ROC-AUC та ECE

Варіант	ROC-AUC	ECE
Онлайн-модель із калібруванням (ізотонічна регресія)	$0,86 \pm 0,07$	$0,08 \pm 0,04$
Онлайн-модель без калібрування	–	$0,18 \pm 0,07$
Проста шкала-проксі (лише пропуски або лише варіативність похибки)	–	–

Табл. 2. Мікробенчмарк латентності оцінювання надійності, μs

Процедура/варіант	Середня латентність, μs
Проста шкала-проксі: лише пропуски	9,29
Проста шкала-проксі: лише варіативність похибки	19,19
Онлайн-модель оцінювання надійності (повний варіант)	≈ 150

Обговорення результатів

Наведені результати підтверджують, що в поточному сценарії корисність шкали надійності визначається двома взаємопов'язаними властивостями: (i) здатністю відокремлювати деградаційні стани та (ii) каліброваністю ймовірнісної інтерпретації. Висока ROC-AUC означає, що скор придатний для ранжування сегментів за рівнем деградації, тоді як низька ECE робить можливим використання числових порогів без повторної «підгонки» під кожен окремий

запуск. Саме каліброваність є тим компонентом, який перетворює скор з діагностичного індикатора на керувальний інтерфейс, придатний для аудитованих політик.

У частині часових характеристик принциповим є те, що метод зберігає мікросекундний порядок затримок. Прості проксі-шкали є швидшими, але повна онлайн-модель демонструє вищу якість за середньої латентності близько $150\mu\text{s}$, що є сумісним із практичними потоковими конвеєрами, де обробка виконується покроково або в малих вікнах. Таким чином, калібрування може розглядатися як «дешева» надбудова, що додає інтерпретованість без радикального збільшення витрат.

Разом із тим межі застосовності визначаються природою проксі-сигналів. Проксі-ознаки деградації не гарантують однаково якісного розділення «надійна / ненадійна» в усіх доменах і для всіх типів деградацій, а компоненти, що використовують статистики похибки, припускають доступність фактичного значення цілі y_i у потоці з допустимою затримкою. У сценаріях із відкладеною розміткою для підтримання коректності оцінювання можуть знадобитися альтернативні проксі або затримані оновлення калібрування. З практичного погляду ці обмеження означають, що модуль надійності потребує експлуатаційного моніторингу та періодичної перевірки узгодженості шкали з реальними подіями деградації, щоб зберегти аудитованість політик керування.

З позиції внутрішньої валідності важливо враховувати, що проксі-сигнали можуть реагувати не лише на деградацію джерела, а й на зміни розподілу даних, викликані концептуальним дрейфом або зміною режиму системи. Це створює потенційну конфундацію «дрейф \leftrightarrow деградація», яку в практиці слід зменшувати через спільний аналіз декількох сигналів і через експлуатаційні контексти (журнали інцидентів, сервісні події, інформація про технічний стан сенсора). З позиції конструктивної валідності є мірою ранжування і не гарантує оптимальності конкретного порога, тоді як ЕСЕ залежить від вибору бінування та від репрезентативності калібрувального сегмента. Тому метрики мають інтерпретуватися як взаємодоповнювальні, а не як взаємозамінні.

З позиції зовнішньої валідності ROC-AUC переносимість проксі-сигналів і калібрування між доменами не є автоматичною. Шкала, калібрована на одному типі деградацій або на одному домені, може втратити узгодженість на іншому через інші механізми пропусків або інші шумові характеристики. Отже, у практичній системі підтримки прийняття рішень модуль надійності слід розглядати як компонент, що потребує періодичної верифікації та (за наявності міток) повторного калібрування.

Висновки

У статті представлено підхід до онлайн-оцінювання надійності модальностей і джерел у поточному аналізі мультимодальних часових рядів із калібруванням ізотонічною регресією. Калібрований варіант демонструє високу здатність відокремлювати деградовані стани (ROC-AUC $0,86 \pm 0,07$) та суттєво покращує каліброваність шкали ($0,08 \pm 0,04$ проти $0,18 \pm 0,07$ без калібрування) за середньої латентності близько $150\mu\text{s}$. Отриманий підхід формує інтерпретований керувальний сигнал, який може бути використаний для стабільного керування компонентами потокової системи підтримки прийняття рішень за деградації модальностей та змінності середовища.

Конфлікт інтересів

Автори декларують, що не мають конфлікту інтересів стосовно цього дослідження, у тому числі фінансового, особистісного характеру, авторства чи іншого характеру, що міг би вплинути на дослідження та його результати, представлені в цій статті.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Дані будуть надані за обґрунтованим запитом.

Подяка

Автори висловлюють щирю подяку рецензентам і редколегії за уважний розгляд рукопису, цінні зауваження й рекомендації.

ЛІТЕРАТУРА

- [1] T. Baltrušaitis, C. Ahuja, L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019. DOI: 10.1109/TPAMI.2018.2798607.
- [2] D. Lahat, T. Adali, C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015. DOI: 10.1109/JPROC.2015.2460697.
- [3] P. Dawid, "Present position and potential developments: Some personal views: Statistical theory: the prequential approach," *Journal of the Royal Statistical Society. Series A (General)*, vol. 147, no. 2, pp. 278–292, 1984. DOI: 10.2307/2981683.
- [4] A. Bifet, J. Montiel, J. Read et al., *Machine Learning for Data Streams with Practical Examples in MOA*, MIT Press, 2018.
- [5] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys*, vol. 46, no. 4, Art. 44, 2014. DOI: 10.1145/2523813.
- [6] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, G. Zhang, "Learning under concept drift: A review," *IEEE Transactions on Knowledge and Data Engineering*,

- vol. 31, no. 12, pp. 2346–2363, 2019. DOI: 10.1109/TKDE.2018.2876857.
- [7] R. S. M. Barros, S. G. T. C. Santos, “An overview and comprehensive comparison of ensembles for concept drift,” *Information Fusion*, vol. 52, pp. 213–244, 2019. DOI: 10.1016/j.inffus.2019.03.006.
- [8] A. Kendall, Y. Gal, “What uncertainties do we need in Bayesian deep learning for computer vision?,” *Advances in Neural Information Processing Systems*, 2017. arXiv:1703.04977.
- [9] Y. Ovadia, E. Fertig, J. Ren et al., “Can you trust your model’s uncertainty? Evaluating predictive uncertainty under dataset shift,” *Advances in Neural Information Processing Systems*, 2019. arXiv:1906.02530.
- [10] E. Hüllermeier, W. Waegeman, “Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods,” *Machine Learning*, vol. 110, pp. 457–506, 2021. DOI: 10.1007/s10994-021-05946-3.
- [11] C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger, “On calibration of modern neural networks,” in *Proc. 34th International Conference on Machine Learning (ICML)*, 2017, pp. 1321–1330.
- [12] J. C. Platt, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” in *Advances in Large Margin Classifiers*, A. J. Smola et al., Eds. MIT Press, 1999, pp. 61–74.
- [13] B. Zadrozny, C. Elkan, “Transforming classifier scores into accurate multiclass probability estimates,” in *Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002, pp. 694–699. DOI: 10.1145/775047.775151.
- [14] A. Niculescu-Mizil, R. Caruana, “Predicting good probabilities with supervised learning,” in *Proc. 22nd International Conference on Machine Learning (ICML)*, 2005, pp. 625–632. DOI: 10.1145/1102351.1102430.
- [15] M. Kull, T. M. Silva Filho, P. Flach, “Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers,” in *Proc. 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 623–631.
- [16] J. Vaicenavicius, D. Widmann, C. Andersson et al., “Evaluating model calibration in classification,” in *Proc. 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019, pp. 3459–3467.
- [17] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. DOI: 10.1016/j.patrec.2005.10.010.

ONLINE RELIABILITY ESTIMATION OF SOURCES IN STREAMING ANALYSIS OF MULTIMODAL TIME SERIES WITH ISOTONIC REGRESSION CALIBRATION

Illia Uzun, Mykhaylo Lobachev

Streaming intelligent decision support systems processing multimodal time series operate under causality constraints and must satisfy requirements of low

latency, bounded computational budgets, and controllable responses to environmental change. A critical practical risk in such pipelines is the temporary degradation of individual sources (missing values, elevated noise, scale shifts), which can masquerade as concept drift and trigger unstable or excessive control actions. This paper considers online estimation of source reliability as a causal probabilistic assessment of being in a non-degraded state and shows that practical control requires a calibrated scale: the output value must be interpretable as the frequency of the “non-degraded” regime under relevant conditions. The proposed approach combines lightweight degradation proxy signals suitable for online computation with isotonic regression calibration, which provides a monotone mapping from scores to correct probabilities. Key experimental results demonstrate ROC-AUC of 0.86 ± 0.07 for the calibrated variant and calibration improvement from ECE of 0.18 ± 0.07 (uncalibrated) to ECE of 0.08 ± 0.04 (calibrated) at acceptable time costs: simple proxy scales have microsecond latencies, while the full online model maintains mean latency of approximately $150 \mu\text{s}$, meeting the needs of streaming pipelines.

Keywords: machine learning, data analysis, information systems, decision support systems, multimodal time series, online calibration, isotonic regression, degradation detection, reliability estimation, streaming data.

REFERENCES

- [1] T. Baltrušaitis, C. Ahuja, L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019. DOI: 10.1109/TPAMI.2018.2798607.
- [2] D. Lahat, T. Adali, C. Jutten, “Multimodal data fusion: an overview of methods, challenges, and prospects,” *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015. DOI: 10.1109/JPROC.2015.2460697.
- [3] P. Dawid, “Present position and potential developments: Some personal views: Statistical theory: the prequential approach,” *Journal of the Royal Statistical Society. Series A (General)*, vol. 147, no. 2, pp. 278–292, 1984. DOI: 10.2307/2981683.
- [4] A. Bifet, J. Montiel, J. Read et al., *Machine Learning for Data Streams with Practical Examples in MOA*, MIT Press, 2018.
- [5] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, “A survey on concept drift adaptation,” *ACM Computing Surveys*, vol. 46, no. 4, Art. 44, 2014. DOI: 10.1145/2523813.
- [6] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, G. Zhang, “Learning under concept drift: A review,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346–2363, 2019. DOI: 10.1109/TKDE.2018.2876857.
- [7] R. S. M. Barros, S. G. T. C. Santos, “An overview and comprehensive comparison of ensembles for concept

- drift,” *Information Fusion*, vol. 52, pp. 213–244, 2019. DOI: 10.1016/j.inffus.2019.03.006.
- [8] A. Kendall, Y. Gal, “What uncertainties do we need in Bayesian deep learning for computer vision?,” *Advances in Neural Information Processing Systems*, 2017. arXiv:1703.04977.
- [9] Y. Ovadia, E. Fertig, J. Ren et al., “Can you trust your model’s uncertainty? Evaluating predictive uncertainty under dataset shift,” *Advances in Neural Information Processing Systems*, 2019. arXiv:1906.02530.
- [10] E. Hüllermeier, W. Waegeman, “Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods,” *Machine Learning*, vol. 110, pp. 457–506, 2021. DOI: 10.1007/s10994-021-05946-3.
- [11] C. Guo, G. Pleiss, Y. Sun, K. Q. Weinberger, “On calibration of modern neural networks,” in *Proc. 34th International Conference on Machine Learning (ICML)*, 2017, pp. 1321–1330.
- [12] J. C. Platt, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” in *Advances in Large Margin Classifiers*, A. J. Smola et al., Eds. MIT Press, 1999, pp. 61–74.
- [13] B. Zadrozny, C. Elkan, “Transforming classifier scores into accurate multiclass probability estimates,” in *Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002, pp. 694–699. DOI: 10.1145/775047.775151.
- [14] A. Niculescu-Mizil, R. Caruana, “Predicting good probabilities with supervised learning,” in *Proc. 22nd International Conference on Machine Learning (ICML)*, 2005, pp. 625–632. DOI: 10.1145/1102351.1102430.
- [15] M. Kull, T. M. Silva Filho, P. Flach, “Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers,” in *Proc. 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 623–631.
- [16] J. Vaicenavicius, D. Widmann, C. Andersson et al., “Evaluating model calibration in classification,” in *Proc. 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019, pp. 3459–3467.
- [17] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. DOI: 10.1016/j.patrec.2005.10.010.

Дата першого надходження статті до видання:

12.02.2026

Дата прийняття статті до друку

після рецензування: 17.03.2026

Дата публікації (оприлюднення) статті:

12.05.2026



Стаття поширюється на умовах ліцензії відкритого доступу CC BY 4.0