DOI: 10.34185/1991-7848.itmm.2025.01.071

ОПТИМІЗАЦІЯ СЕРВЕРНОГО-КЕШУВАННЯ ЗА ДОПОМОГОЮ АДАПТИВНОГО TTL НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Сімакін С.К.[1], Божуха Л.М.[2]

[1]Дніпровський національний університет імені Олеся Гончара, аспірант,Україна

[2]Дніпровський національний університет імені Олеся Гончара, кандидат фізико-математичних наук, доцент, Україна

Анотація. *Статичне визначення часу життя (TTL) кешованих даних на сервері часто призводить до неефективного використання ресурсів або подачі застарілої інформації. Запропоновано підхід до оптимізації кешування шляхом динамічного визначення TTL за допомогою машинного навчання. Розглядаються два основні методи: пакетне навчання моделі (наприклад, градієнтний бустинг) на основі історичних даних про частоту змін джерела, патерни доступу та характеристики даних; та онлайн-навчання, зокрема з використанням навчання з підкріпленням (RL), яке дозволяє системі безперервно адаптуватися до поточних умов. RL-агент навчається обирати оптимальний TTL, аналізуючи позитивні (влучання в кеш) та негативні (промахи, застарілі дані) результати своїх дій. Очікується, що адаптивний TTL підвищить cache hit rate, зменшить затримки та навантаження на базу даних, забезпечуючи при цьому вищу актуальність даних.*

Ключові слова: *кешування, time-to-live (TTL), оптимізація продуктивності, машинне навчання, адаптивне керування, вебдодатки, серверне кешування, навчання з підкріпленням.*

Серверне кешування є невід'ємною частиною сучасних вебдодатків, дозволяючи підвищити швидкість відгуку та зменшити навантаження на джерела даних. Ключовим параметром є час життя (TTL) кешованих даних. Традиційний підхід із фіксованим TTL часто є неефективним: занадто короткий TTL – знижує користь від кешу, а занадто довгий – підвищує ризик подачі застарілих даних. Налаштування оптимального статичного TTL є складною задачею.

Для вирішення цієї проблеми пропонується концепція адаптивного TTL, що використовує машинне навчання (ML) для динамічного визначення часу життя кешу. Метою цього дослідження є представлення та аналіз ML-підходів для реалізації адаптивного TTL у системах серверного кешування [1].

Одним з недоліків статичного TTL є фіксований час життя кешу, який погано адаптується до динамічної природи вебданих. Основними проблемами цього підходу є неефективне використання кешу (занизький TTL), ризик подачі неактуальних даних (зависокий TTL), складність ручного налаштування та підтримки оптимальних значень для різних типів контенту.

Для вирішення цієї проблеми можна використовувати адаптивний TTL на основі машинного навчання. Ідея полягає в автоматичному визначенні TTL для кожного елемента кешу за допомогою ML-моделей, що аналізують характеристики даних та патерни їх використання.

Підходи до реалізації. Розглянуто пакетне навчання (Batch Learning), що має такі властивості, як:

- модель (наприклад, градієнтний бустинг) навчається офлайн на історичних даних;
- вхідні дані (ознаки): частота доступу до елемента, частота його змін у джерелі, тип даних, поточне навантаження системи тощо;
- вихідні дані: прогнозоване значення TTL;
- використання моделі: офлайн-прогноз, TTL розраховуються заздалегідь для багатьох ключів і зберігаються. При кешуванні береться готове значення (швидко). Онлайн-прогноз, TTL розраховується моделлю безпосередньо в момент кешування (точніше враховує поточний стан, але додає затримку).

Запропоновано на розгляд, як альтернативу, - онлайн-навчання (Online Learning), зокрема Навчання з Підкріпленням (RL) з наступними властивостями:

- система навчається безперервно, адаптуючись до змін у реальному часі;
- RL-підхід: Агент (система вибору TTL) виконує дію (встановлює певний TTL) в залежності від поточного стану (ознаки кешованого елемента та системи). Потім агент отримує зворотний зв'язок – "винагороду" (позитивну за вдалі дії, як-от cache hit) або "штраф" (негативний за невдалі дії, як-от cache miss або подачу застарілих даних). На основі цього зворотного зв'язку агент коригує свою стратегію вибору TTL, прагнучи максимізувати позитивні результати в довгостроковій перспективі.

Реалізація потребує компонентів для збору даних (логи доступу, зміни джерел), обробки ознак, роботи ML-моделі (як сервісу або вбудованої логіки) та інтеграції з існуючою системою кешування.

Перевагами на цьому етапі адаптивного підходу є потенційне підвищення ефективності кешування (hit rate), краща актуальність даних та автоматизація

налаштувань. Але в порівнянні з онлайн-навчанням недоліком є збільшення складності системи, додаткові витрати ресурсів на ML, необхідність ретельного проєктування та налаштування.

Висновки. Адаптивний підхід до визначення TTL у серверному кешуванні на основі машинного навчання є перспективною альтернативою статичним методам. Розглянуті підходи (пакетне та онлайн/RL навчання) дозволяють динамічно керувати часом життя кешу, потенційно покращуючи баланс між продуктивністю та актуальністю даних. Хоча реалізація таких систем є складнішою, очікувані переваги виправдовують подальші дослідження, розробку та експериментальну перевірку ефективності цих методів у реальних умовах експлуатації вебдодатків.

ЛІТЕРАТУРА

1. Мельник А. В. Оптимізація ефективності кешування в інформаційно-орієнтованих мережах за допомогою аналізу трафіку і використання адаптивних алгоритмів. Комп'ютерні системи та мережні технології. – 2023. С. 115–116.

OPTIMIZATION OF SERVER CACHING USING ADAPTIVE TTL BASED ON MACHINE LEARNING

Simakin S., Bozhukha L.

Abstract. *Static definition of time-to-live (TTL) for cached server data often leads to inefficient resource usage or serving stale information. An approach for caching optimization through dynamic TTL determination using machine learning is proposed. Two main methods are considered: batch training of a model (e.g., gradient boosting) based on historical data about source change frequency, access patterns, and data characteristics; and online learning, particularly using reinforcement learning (RL), which allows the system to continuously adapt to current conditions. The RL agent learns to select the optimal TTL by analyzing positive (cache hits) and negative (misses, stale data) outcomes of its actions. Adaptive TTL is expected to increase the cache hit rate, reduce latency and database load, while ensuring higher data freshness.*

Keywords: *caching, time-to-live (TTL), performance optimization, machine learning, adaptive control, web applications, server caching, reinforcement learning.*

REFERENCE

1. Melnyk A. V. Optymizatsiia efektyvnosti keshuvannia v informatsiino-oriientovanykh merezhakh za dopomohoiu analizu trafiku i vykorystannia adaptyvnykh alhorytmiv. Komp'iuterni systemy ta merezhni tekhnolohii. – 2023. P. 115–116. [in Ukrainian].