

UDC 044.9. 629.7.05

DOI: 10.18372/2073-4751.71.17002

Радченко К.О.,

orcid.org/0000-0002-1282-6307

## ОСОБЛИВОСТІ ПРОГНОЗУВАННЯ РІВНЯ ВЕБТРАФІКУ У КОМП'ЮТЕРНИХ МЕРЕЖАХ ЗАГАЛЬНОГО ПРИЗНАЧЕННЯ

Київський політехнічний інститут ім. Ігоря Сікорського

radchenko.kostiantyn@iit.kpi.ua

### Вступ

Будемо розглядати веб (WWW) як великомасштабну розподілену інформаційну систему, з вираженими особливостями просторової і часової локалізації. Потік вебпосилань на практиці залежить у часі і просторі від кешування, попередньої вибірки та системи поширення документів.

Часова локалізація передбачає, що документи, до яких нещодавно зверталися, з більшою ймовірністю будуть згадані в найближчому майбутньому. Ця властивість привела дослідників до пропозиції створити стек всіх об'єктів сервера, впорядкованих за давністю використання [1]. Просторова локалізація оперує з тим, що загальна кількість унікальних послідовностей різної довжини, що спостерігаються у потоці вебпосилань, набагато менша від очікуваного у випадково обраному вебпосиланні [5,14].

Локалізація є одним з чинників, які впливають на базовий байтовий вебтрафік, який виникає в результаті передачі різних документів HTTP, було показано, що вебтрафік дуже мінливий у широкому діапазоні відрізків часу [8].

З іншого боку, вебтрафік може бути описаний статистично з використанням поняття самоподібності, яке може бути застосоване до пульсацій вебтрафіку глобальної і локальної мережі [14]. Через кореляції, які є присутні при масштабуванні, стаціонарні самоподібні процеси мають властивість далекодії. Залежні часові ряди з великим діапазоном зазвичай мають сплески, тобто значення вище чи нижче середнього, які виникають у всіх часових рамках. Отже, незалежно від масштабування ряду ми можемо спостерігати сплески.

Самоподібність у вебтрафіку можна пояснити на основі ефектів кешування користувачьких уподобань при передачі, ефекта клієнтського «часу на роздуми» і накладання багатьох таких передач у локальній області розподілу розмірів вебдокументів [5].

Така масштабно-інваріантна мінливість різко контрастує з традиційними моделями вебтрафіку близькодії, описаними пуассонівськими чи марківськими процесами, які мають характерну довжину сплеска з тенденцією до згладжування шляхом усереднення у достатньо тривалому часовому масштабі. Часовий ряд із довгостроковою залежністю має набагато складніші закономірності своєї поведінки, аніж ряд з короткостроковою залежністю. Враховуючи складні аспекти довгострокової залежності, самоподібність була доведена у [14], і важливість прогнозу довгострокової залежності починає відзначатись у таких дослідженнях, як [5,8], які показують, що втрата пакетів та поведінка затримки радикально відрізняється при моделюванні з використанням даних реального трафіку, а не традиційних мережевих досліджень.

Зрозуміло, що ефективне прогнозування, здатне наперед з достатньою точністю надавати оцінки попиту на вебтрафік, могло би значно допомогти оптимізації комп'ютерної системи та мережевих ресурсів. Прогнозування кількості переданих байтів через вебтрафік дозволило б передбачити очікувану продуктивність для проектування мережі та планування пропускної спроможності комп'ютерної мережі загального призначення. Більше того, довгострокова характеристика вебтрафіку також дозволить нам генерувати синтетичні потоки для моделювання вебтрафіку,

що має значення для мережевих операторів.

### **Аналіз досліджень та публікацій**

Дослідження, які спрямовані на підвищення точності прогнозування вебтрафіку мають у своєму складі ряд методів основаних на штучних нейронних мережах, в тому числі рекурентних, оптимізаційних методах лінійного та не лінійного програмування тощо [1-3]. Ми показуємо, як використання конекціоністських моделей, зокрема нейронних мереж, може допомогти у збиранні корисної інформації у різних часових масштабах. Нейронні мережі зробили вагомий внесок, особливо у сфері класифікації та передбачення часових рядів [2,13]. Однак було показано, що алгоритми навчання нейронних мереж з градієнтним спуском можуть дуже погано працювати із даними, які мають довгострокову залежність [4], тобто з такими даними, для яких бажаний результат залежить від вхідних даних, які були представлені значно раніше. Проблема може бути зменшена за рахунок використання рекурентних нейронних мереж (RNN), зокрема LSTM, які мають потужні репрезентативні можливості. Ці моделі наділені часовим виміром за рахунок вбудовування авторегресійних фільтрів у повністю рекурентну нейронну структуру. Градієнтний спуск у цій складнішій нейронній моделі, що допускає приховані стани, може бути більш ефективним.

На додаток до цього ми припускаємо, що мережева архітектура, що працює в кількох масштабах, допомагає вирішити проблеми різкого зменшення точності прогнозування. Наша стратегія прогнозування полягає у розбитті завдання прогнозування на елементарні, більш прості, завдання. Для цього ми використаємо дискретне вейвлет-перетворення, призначене для виявлення особливостей мінливості вебтрафіку, які потім обробляються нейронною мережею. Як і перетворення Фур'є, вейвлет-перетворення (ВП) є універсальним інструментом, що дозволяє краще обробляти дані. Воно ефективно

використовується для стиснення зображень, видалення шуму, виявлення об'єктів та великомасштабного структурного аналізу [4,6,10-12]. Вейвлет-перетворення є надлишковим перетворенням (без проріджування даних). Суть ВП полягає у розкладі вхідних даних на деталізовані сигнали та залишковий сигнал, так що вихідний сигнал може бути виражений як адитивна комбінація вейвлет-коефіцієнтів при різних рівнях роздільної здатності. Останні забезпечують декомпозицію сигналу або часового ряду, так що можуть бути помічені короткочасові особливості. Таким чином рекурентна нейронна мережа спрацьовує кожен раз, щоб надати оцінки вейвлетів на заданому часовому кроці, які будуть об'єднані для формування прогнозу наступного часового кроку. Інші методи прогнозування можуть використовуватися спільно і вибиратися найкращі на кожному рівні доступу.

Цей підхід успішно був застосований до часового ряду 5-хвилинних середніх значень вебтрафіку. Ми зібрали та проаналізували журнал вебтрафіку однієї з регіональних аварійно-диспетчерських служб в Україні і отримали повну картину локального веб-трафіку з типовими особливостями поведінки досліджуваного вебсервера протягом одного календарного року. Наша робота заснована на даних, зібраних на рівні вимірювання величини вебтрафіка і включають записи про його значення у кожен конкретний момент часу. Наша увага не спрямована на встановлення причин часової локальності чи самоподібності. Натомість ми зосередимося на розрахунку прогнозування у годинному та хвилинному масштабі після того, як коротко показали, що ряди мають властивість довгострокової залежності.

### **Мета**

Дослідити архітектуру нейронної мережі та особливості використовуваного вейвлет-перетворення, що використовується для моделювання. Побудувати власну модель та провести її тестування з метою визначення ступеню випадкових відхилень вихідних параметрів. Провести

апроксимацію отриманої кривої за даною моделлю та обговорити результати.

### Основна частина

З метою прогнозування рівня вебтрафіку змодельовано нейронну мережу вітряних турбін, яка повинна мати можливість зберігати та оновлювати контекстну інформацію, що відповідає довготривалим часовим інтервалам. Загальний підхід передбачає явне включення затримок до рівняння нейронної системи. У такій системі мультиплікативні взаємодії скалярів замінюються операціями часової згортки, що вимагають вагової матриці, яка залежить від часу [2]. Загальна формула для мереж із затримкою з неперервним часом є рівняння виду:

$$\begin{aligned} T \frac{dv(t)}{dt} &= g(u(t)) + i(t), \\ u(t) &= \int_0^t W(t-s)v(s)ds, \end{aligned} \quad (1)$$

де  $N$  – загальна кількість проміжків,  $u(t)$  і  $v(t)$  –  $N$ -вимірні векторна активність і внутрішні входи на проміжках в момент часу  $t$ ,  $W$  –  $N \times N$  матриця синаптичних взаємодій, що залежить від часу,  $T$  – діагональна матриця зручно обраної позитивної постійної часу, а  $i(t)$  – вхідний вектор. Сигмоїдальна функція  $g(h) = \frac{1}{1+e^{-\beta h}}$  є звичайним вибором для нелінійної функції активації.

Параметр  $\beta$  керує крутизною логістичної функції і довільно встановлюється рівним  $1/2$ . Надалі зауважимо, що  $g(u_k) = [(u_k^1), \dots, (u_k^n)]^T$ .

Оскільки більшість реалізацій нейронних мереж формулюються для дискретного часу, ми можемо легко перетворити (1) на набір різницевих рівнянь першого порядку, використовуючи  $\frac{du}{dt} = \frac{u(t+\Delta t) - u(t)}{\Delta t}$  разом з дельта-функцією ядра  $W(t) = \sum_{d=0}^D W^d \delta(t-d)$  [2,12]. Приймаючи  $\Delta t = 1$ , остаточно отримуємо дискретну версію (1). Що є гарним наближенням за умови, що одиниця часу, що відповідає  $\Delta t = 1$ , досить мала по відношенню до динаміки стану та параметрів. Питання, наскільки це дискретне представлення є точним, виходить за рамки цієї статті.

Тепер розглянемо результуючу мережу з дискретним часом. Нехай матричний елемент  $w_{ij}^d$  позначає силу з'єднання, або зв'язок, від  $d$ -ї ланки затримки, що з'єднує вузол  $i$  з вузлом  $j$ , а  $D$  – максимальну затримку з'єднання, тобто порядок фільтра,  $v_k$  являє собою вектор активності сигналів, коли час дорівнює  $k$ . Ми не допускаємо з'єднань з вхідними сигналами, тобто  $w_{ij}^d = 0$ , якщо  $j$  є вхідним сигналом, і встановлюємо  $i_k(j) = 0$ , якщо сигнал  $j$  не вхідний. Вихідні сигнали мають цільові значення  $d_k$ . Крім того, вхідне зміщення для кожного сигналу досягається шляхом фіксації активації конкретного сигналу на одиницю. Для зручності запису ми не розрізнятимемо вхідні, приховані, вихідні та зміщені сигнали в наступних формулах.

Також, ваги для  $d$ -их з'єднань із затримкою блоку  $j$  на ітерації  $k$ , задаються вектором  $w_k^d(j) = (w_k^d(1, j), \dots, w_k^d(N, j))^T$ . Так само матриця  $W_k^d = [w_k^d(1), w_k^d(2), \dots, w_k^d(N)]$  розмірністю  $N \times N$  містить всі адаптивні ваги затримки  $d \leq D$  на ітерації  $k$ .

Ми визначаємо відповідні параметри системи на кожному кроці  $k$  шляхом визначення наступної системи пов'язаних диференціальних рівнянь:

$$\begin{aligned} T \frac{dv_k(t)}{dt} &= -v_k(t) + g(u_k(t)) + I_k, \\ u_k(t) &= \sum_{d=1}^D (W_k^d)^T v_{k-d} + (W_k^0)^T v_k(t), \end{aligned} \quad (2)$$

Зовсім не обов'язково, що рівняння (2) завжди має стійкий рівноважний розв'язок, і не матиме можливої коливальної чи хаотичної поведінки [9]. Ми не будемо розглядати це надалі і припустимо існування принаймні одного стабільного атрактора з нерухомою точкою. Тому, якщо такий розв'язок існує для системи (2), то легко бачити, що це підтверджує рівняння:

$$v_k = g(\sum_{d=0}^D (W_k^d)^T v_{k-d}) + I_k. \quad (3)$$

Зручний для реалізації алгоритм навчання цієї моделі запропонований в [2]. При чому на кожному часовому кроці потрібно порядку  $O(D^2 N^2)$  операцій налаштування параметрів мережі.

Тепер у нас є необхідна теоретична база стосовно нейронних мереж [4,22].

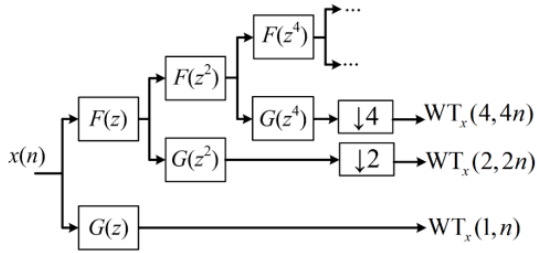


Рис. 1. Реалізація дискретного вейвлет-перетворення у вигляді деревоподібної структури.

Розглянемо апроксимацію часового ряду, використовуючи дискретне вейвлет-перетворення (ДВП) часових рядів як вхідний вектор значень дискретної вибірки. Вейвлети можна розглядати як банки фільтрів нижніх і верхніх частот, що перетворюють вектор даних не змінюючи довжини цього вектора [2].

Вейвлет-перетворення для дискретних даних буде забезпечуватися алгоритмом *à trous* (з дірками) (рис. 1) [7] з надлишковим перетворенням, тобто без провіджування. На рис. 1 через  $G(z)$  позначено  $z$ -перетворення базисного вейвлета, а через  $F(z)$  – передатну функцію інтерполяційного фільтра, що відповідає масштабній функції.

Алгоритм можна описати в такий спосіб. Спочатку виконайте послідовні згортки з дискретним фільтром низьких частот  $h$ :

$$c_{i+1}(k) = \sum_{l=-\infty}^{+\infty} h(l)c_i(k + 2^i l). \quad (4)$$

Нульовий коефіцієнт представляє вихідний ряд:  $c_0(t) = x(t)$ . Збільшення відстаней між точками вибірки ( $2^i l$ ) пояснює, чому цей метод отримав назву «*à trous*». Фільтр низьких частот  $h$ , який ми використовували, був сплайн  $B_3 = (\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$ . Він є компактним та точково-симетричним. Оскільки час є асиметричною змінною в принципі, ми використовуємо функцію симетричного згладжування.

Також візьмемо різницю між послідовними згладженими версіями сигналу, щоб отримати вейвлет-коефіцієнти  $w_i$ :

$$w_i(k) = c_{i-1}(k) - c_i(k) \quad (5)$$

Останні забезпечують більш деталізований сигнал, який, як ми сподіваємося, на практиці фіксуватиме невеличкі деталі важливі для дослідження. Щоб така характеристика не втрачала всіх особливостей своєї поведінки, має бути можливість відновити вихідний вектор даних компонентів вейвлета. Крім того, залишковий член  $c_p$ , що представляє "фонову" інформацію даних, зрештою додається до вейвлет-коефіцієнтів.

Множина  $W = \{w_1, w_2, \dots, w_p, c_p\}$  представляє вейвлет-перетворення даних глибиною  $p$ . Зрозуміло, що при прогнозуванні необхідно приділяти пильну увагу межах. Обробка меж може здійснюватись по-різному. Ми використовували інтуїтивний підхід, взявши (для часового ряду розміру  $N$ )

$$c(N + k) = c(N - k).$$

Вейвлет-розклад часового ряду через вейвлет-коефіцієнти визначається виразом:

$$x(t) = c_p(t) + \sum_{i=1}^p w_i(t). \quad (6)$$

Легко бачити, що обчислювальна складність описаного вище алгоритму становить  $O(n)$  для  $n$ -значних вхідних даних, а обсяг пам'яті дорівнює  $O(n^2)$ .

Тепер поєднаємо нейронно-вейвлетний підхід. Метод полягає в ієрархічному застосуванні фільтра низьких частот до часового ряду, спочатку до вихідного вектора даних  $c_0$ , потім до згладженого вектора  $c_i$  і так далі, поки результуючий вектор  $c_p$  не буде настільки згладженим, що його прогнозування на один крок уперед стане тривіальним (напр. лінійним чи його копією). Для адитивності розкладу вейвлет-перетворення ми, звичайно, маємо враховувати вейвлет-коефіцієнти, задані згорткою даних та фільтром низьких частот. Ми можемо використати послідовні згортки типу ковзної середньої для  $N$  віддалених точок. У момент часу  $t$  ми маємо спостереження точок  $x(t), x(t-1), \dots, x(1)$ , і ми шукаємо значення  $x(t+1)$ . Зупинивши розклад на  $p$ -му рівні, ми отримали б масив вейвлет-коефіцієнтів, до якого треба додати нев'язку, що позначається  $\{W(i, j)\}$  для  $i = t, t-1, \dots, 1$  та  $j = 1, 2, \dots, p+1$ . Це

адитивний розклад вхідного сигналу, тому ми маємо

$$x(i) = \sum_{j=1}^{p+1} W(i, j), \forall i = 1, 2, \dots, t. \quad (7)$$

Рівняння (7) також забезпечує формулу реконструкції вихідного часового ряду. Ідея полягає у тому, щоб прогнозувати значення кожного  $W(i, j)$ . Навчання прогностичних вибірок здійснюється класичним методом, тобто. на навчальній підмножині даних.

Оскільки реконструкція є адитивною, ми можемо об'єднувати прогнозування також адитивним шляхом. Однак, ми можемо прийняти гібридну стратегію щодо об'єднання даних для отримання загального прогнозу. Тобто, ми можемо протестувати ряд передбачень для короткої та довгої пам'яті на кожному рівні дозволу та використати комбінацію, яка має мінімальні похибки.

Інтернет є великомасштабною розподіленою інформаційною системою, заснованою на архітектурі клієнт-сервер. Таким чином, робоче навантаження на вебсервер, складається з низки запитів, що походять від багатьох клієнтів. Ми вивчили журнали доступу для досліджуваного сервера, а саме вебсервера однієї з регіональних аварійно-диспетчерських служб в Україні (з червня 2021 р. до вересня 2022 р.). Журнали містять інформацію для кожного запиту, обробленого сервером, включаючи ім'я хоста, з якого виходить запит, позначку часу запиту, ім'я запрошеного файлу об'єкта і розмір відповіді в байтах.

Зважаючи на розрідженість даних від досліджуваного вебсервера з 5-хвилинними інтервалами, ми побудували перший набір даних, заснований на нормалізованому по діапазону середньогодинному рівню вебтрафіку, тобто запитах на передачу, задоволених сервером. Перевірка даних показала різні режими в часовому ряді через різке підвищення вебтрафіку щодня у робочий час, крім вихідних днів та фонове навантаження на вебсервер протягом неробочих днів і годин. Тому ми вирішили розділити часовий ряд на підсерії тижневої тривалості і використовувати середніх 5-хвилинні значення для експериментів.

Графік на рис. 2 являє собою частину представлення часового ряду вебтрафіку протягом листопада 2021р. з використанням 1-годинних інтервалів. Явно видно добовий цикл попиту мережі, а повсякденна активність демонструє помітний фон.

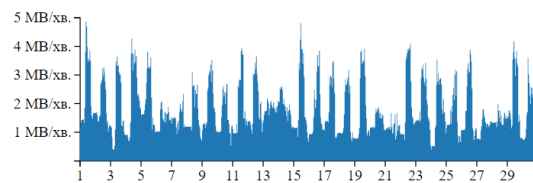


Рис. 2. Рівень вебтрафіку досліджуваного вебсервера протягом місяця.

Із графіка чітко видно періодичність добових коливань вебтрафіку, менш вираженою є тижнева періодична компонента.

На цьому етапі ми можемо запитати себе, яка частота найкраще підходить для вимірювання трафіку – кожену секунду, кожні півхвилини тощо – пам'ятаючи про те, що погодинна частота маскує дрібніші зміни масштабу, тоді як надмірна частота може призвести до небажаних шумових коливань. Однак краще мати багато даних, а не мало, оскільки дані завжди можна агрегувати зменшуючи варіації.

У статті ми не концентруємося на з'ясуванні причин самоподібності вебтрафіка, і тому ми використовуємо лише простий графічний метод для перевірки самоподібності вебпосилань [5,14].

Цей метод – залежність від часу, заснована на дисперсії самоподібного ряду, що повільно загасає. Самоподібний часовий ряд має ту властивість, що при агрегуванні новий ряд має ту ж саму функцію автокореляції, що і попередній. Тобто для заданого стаціонарного часового ряду  $X_t$ , для  $t = 1, 2, \dots, m$ -агрегований член  $X_t^m$  отримується шляхом підсумовування по блоках розміру  $m$ , що не перетинаються. Тоді, якщо  $X_t$  самоподібний, автокореляція  $E[(X_t - m_X)(X_{t+k} - m_X)]$  має той самий розподіл для всіх  $X_t^m$ , як і вихідний ряд. В результаті автомодельний процес проявляє довгострокову залежність. Його автокореляційна функція  $r(k) \sim k^{-\beta}$  при  $k \rightarrow \infty$ , де  $\beta < 1$ . Таким чином,

автокореляційна функція такого процесу згасає гіперболічно, на відміну від повільнішого експоненційного згасання, що демонструється традиційними моделями. Коли дисперсія  $X_t^m$  відображається залежно від  $m$  на логарифмічному графіку, пряма лінія з нахилом  $\beta$ , що перевищує  $-1$ , вказує на самоподібність, а параметр Херста  $H = 1 + \beta/2$  визначається в термінах асимптотичної поведінки масштабованого діапазону як функції відрізка часу часового ряду наступним чином:

$$M\left[\frac{R(n)}{S(n)}\right] = Cn^H, n \rightarrow \infty, \quad (8)$$

де  $R(n)$  – розмах накопичених відхилень перших  $n$  значень від середнього ряду,  $S(n)$  – стандартне відхилення,  $M[x]$  – математичне сподівання,  $n$  – кількість точок на відрізку часового ряду),  $C$  – постійний параметр.

Для самоподібного ряду  $1/2 < H < 1$ , і при  $H \rightarrow 1$  ступінь самоподібності ряду зростає.

Крім того, використовувані розподіли цілком можуть бути охарактеризовані за допомогою логнормальних розподілів. Однак логнормальні розподіли не мають необмеженої дисперсії і, отже, не мають важких "хвостів" у строгому значенні. Щоб перевірити, чи наші набори даних відповідають характеристикам нескінченної дисперсії важких "хвостів", ми використали простий тест, заснований на центральній граничній теоремі (ЦПТ), який стверджує, що сума незалежних та однаково розподілених вибірок з будь-якого розподілу з обмеженою дисперсією мають тенденцію до логнормальних розподілів. Для цього ми перевіряємо поведінку "хвоста" агрегованих наборів даних, використовуючи графіки логарифмічного додаткового кумулятивного розподілу (LLCD). Це графіки додаткових кумулятивних розподілів  $\hat{F}(t) = 1 - F(x) = P[X > x]$  на логарифмічних осях. Побудовані таким чином графіки розподілів з важкими "хвостами" мають лінійний вигляд. Для наборів даних з обмеженою дисперсією нахил все більше зменшуватиметься в міру збільшення  $m$ , відображаючи нормальний розподіл, що

лежить в основі. Для наборів даних із необмеженою дисперсією нахил залишатиметься приблизно постійним із збільшенням  $m$ .

Результат, який характеризує залежність між співвідношенням  $\frac{R(n)}{S(n)}$  та вихідними добовими коливаннями вебтрафіка у відповідності до виразу (8) представлений на рис. 3. Залежність дисперсії-часу середньогодинної кількості сигналів, оцінюється з використанням регресії, яка дає параметр Херста в середньому рівний  $H=0.718$ , що вказує на наявну самоподібність ряду [5,14].

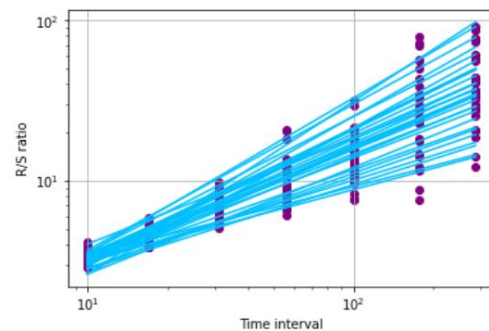


Рис. 3. Залежність параметра Херста від добових коливань вебтрафіка.

Як зазначалося вище, вейвлет-перетворення було застосовано до часового ряду середньогодинних значень рівня вебтрафіка. Хоча було пораховано 5 рівнів вейвлета, останній був відкинтий, тому що ми виявили, що він надто повільно змінюється. Дійсно, чим вищий порядок глибини розрахунку, тим гладкіша крива і, отже, тим менше інформації можна отримати через нейронну мережу від отриманих коефіцієнтів. У свою чергу, здатність мережі збирати інформацію про часові ряди з вищим порядком роздільної здатності дуже швидко погіршується.

Чотири рівні вейвлету, що залишилися, були усічені відповідно до плюс або мінус 100, 80, 50 і 40 і лінійно масштабовані, щоб потрапити в одиничний інтервал. Хоча вихідний нейрон був лінійним і, отже, необмеженим, ми обмежили дані до  $n=400$ , а потім лінійно масштабували кількість влучень, щоб залишатися приблизно в діапазоні  $[0;1]$ . Скорочення кількості

збігів було необхідно, щоб зменшити мінімальну та максимальну величину до розумного значення, оскільки в іншому випадку дуже великі значення розбавили б варіації даних.

Кількісні показники ефективності виражаються у термінах помилок прогнозу в межах 1 крока уперед. Точність визначалась нормалізованою середньоквадратичною помилкою (NMSE):

$$NMSE = \frac{1}{\sigma^2 N} \sum_{k=1}^N (x(k) - \hat{x}(k))^2, \quad (9)$$

де  $x(k)$  – вихідні значення послідовності,  $\hat{x}(k)$  – прогнозовані значення, дбачення, і  $\sigma^2$  – дисперсія вихідної послідовності щодо прогнозованої. Таким чином, значення  $NMSE = 1$  відповідає прогнозуванню безумовного середнього.

Була використана рекурентна архітектура; вейвлети були представлені як зовнішня інформація для мережі. Було використано набір із 2016 кортежів спостереження рівня вебтрафіка за тиждень без пропущених значень. Очевидно, що дані демонструють добові періоди. Непостійні добові коливання величини були найскладнішою характеристикою ряду, яку нейронна мережа могла вивчити. Перші 1728 кортежів використовувалися для навчальних цілей, а прогнозування проводилися для останніх 288 кортежів. Похибка передбачення відстежується на утриманих значеннях, весь наступний тиждень.

Після кількох випробувань була обрана нейронна архітектура LSTM. Використану RNN можна розглядати як трирівневу мережу з 5 x 5 x 1 вузлами з повністю зв'язаними прихованими вузлами з'єднань. Отримана мережа мала 64 посилення, включаючи зміщення. Вибір цих розмірів відбувався здебільшого евристичними методами. Вхідний рівень діє як банк лінійних фільтрів, а модель має коротку довготривалу пам'ять про свої приховані стани.

Мережа, навчена виводити наступне значення до повної збіжності, показала дуже хороші результати (рис. 4) на навчальному наборі, незважаючи на шумові варіації ряду і обмежену кількість

параметрів, що налаштовуються. Навпаки, результуюча нормалізована середньоквадратична помилка для однокрокових прогнозів щодо набору прогнозів становить 0,9, тобто незначна. Годинна агрегація рівня вебтрафіку сильно згладила варіації на менших часових масштабах.

Враховуючи попередній аналіз – зменшимо відрізок прогнозування і сконцентруємось на прогнозах на одну годину вперед відлічуючи від поточного часу доби. Збільшення частоти дискретизації має дві переваги: у модель надходять більш актуальні дані та значно збільшено загальний обсяг даних, доступних для навчання. Дані належать до одного місяця журналів логування особливостей вебтрафіка.

Така сама модель використовувалася для прогнозування роздільної здатності кожного вейвлета. Вибір відповідної кількості ланок, вузлів і тощо в основному визначався емпірично. Як правило, ефективна архітектура забезпечує оптимальний баланс між продуктивністю та складністю.

Обрана модель складається з 5 вхідних блоків, 5 прихованих блоків і одного вихідного блоку. У кожному мережу вводилося значення попереднього ряду для ряду  $W(t, j)$  у певний момент часу  $t$  щоб забезпечити оцінку наступного значення для  $W(t+1, j)$ . Навчальні кортежі склалися з апріорних значень ряду. Хоча використовувалася та ж сама тришарова RNN, що складається з 1 вхідного блоку, затримки з'єднань, що живлять приховані блоки, значення 5 вейвлет-коефіцієнтів  $w_i$ , були змінені, щоб врахувати затухання зміни поведінки  $w_i$  у міру того, як  $i$  зростало від 1 до 4. Чим вище порядок шкали роздільної здатності, тим плавніше крива і більше часове вікно. Затримки виражаються в одиницях часу. Чим більше значення, тим довше затримка передачі між блоками  $i$ , отже, глибше пам'ять. Часове вікно охоплює період тривалістю до 45 хвилин. Зверніть увагу, що максимальна затримка та крок затримки поступово збільшуються.

Такий механізм прогнозування запустився незалежно 5 разів. Як тільки були

отримані одновимірні прогнози для ряду  $W(t,j)$  для  $j$  від 1 до 5, помилка згодом порівнювалася з іншими показниками ефективності і помилкою на основі авторегресійного прогнозу. Ця помилка копіювання зменшувалася при зростанні коефіцієнта. Зверніть увагу, що останні набори вейвлет-коефіцієнтів є наборами, що змінюються більш гладко, отриманими в результаті вейвлет-розкладу.

Нормалізована середньоквадратична помилка для однокрокових рекомбінованих прогнозів RNN за набором прогнозів становить 0,70. Очевидно, що здатність мережі фіксувати динамічну поведінку при зростанні глибини роздільної здатності погіршується дуже швидко. Чим вищий порядок шкали роздільної здатності, тим гладкіша крива і, отже, тим менше інформації може отримати мережа. З цієї причини ми зупинили процес вейвлет-розкладу на 5-му порядку.

Щоб оцінити практичне значення нашого методу, ми використали ту ж саму архітектуру, що і при тижневому прогнозуванні рівня вебтрафіку. Було отримано в результаті  $NMSE=0,85$ . Було з'ясовано, що поєднання RNN із вейвлет-розкладом забезпечує підвищення точності прогнозування.

Розглянемо коефіцієнт взаємної кореляції:

$$p_{ij} = \frac{E[(X_i - m_i)(X_j - m_j)]}{\sigma_i \sigma_j} \quad (10)$$

і визначимо коефіцієнт кореляції:

$$r_{ij} = \frac{E[(\hat{X}_i - X_i)(\hat{X}_j - X_j)]}{\sigma_i \sigma_j} \quad (11)$$

Остаточню NMSE легко виводиться через граничний прогноз NMSE за допомогою

$$NMSE = \frac{\sum_j \sigma_j NMSE_{j+2} \sum_{i>j} r_{ij} \sigma_i \sigma_j}{\sum_j \sigma_j + 2 \sum_{i>j} p_{ij} \sigma_i \sigma_j} \quad (12)$$

Продуктивність передбачення досить швидко погіршується при поєднанні граничних прогнозів. Наприклад, коли рекомбінуються тільки  $w_5$  та  $w_4$ , ми отримуємо  $NMSE=0,13$ ; коли ми додаємо ще  $w_3$ , ми отримуємо  $NMSE=0,24$ . Коли додається і  $w_2$ , тоді ми одержуємо  $NMSE=0,38$ . Зрештою, коли всі оцінки об'єднані, ми отримуємо  $NMSE=0,70$ .

Зі спостережень, наведених вище, ми підкреслюємо, що серії вейвлетів не є незалежними один від одного, так само як і їх прогнози на окремих коефіцієнтах. Коли підсумовуються кілька рядів з різними значеннями NMSE, можна було б очікувати, що результуючий ряд матиме NMSE, що лежить десь між максимальним та мінімальним. Насправді це не так.

Коли всі оцінки вейвлетів рекомбінуються, результуюче значення NMSE становить 0,7, що значно вище, ніж найвища гранична NMSE для вейвлетів, тобто 0,37 для ряду з урахуванням  $w_1$ . Ця значна помилка виникає через сильну взаємну кореляцію помилок прогнозу, яка не була розглянута тут.

Незважаючи на досить відносну точність остаточної оцінки, можна зробити позитивні твердження щодо варіацій на одну годину вперед: у 80% випадків варіація нейронного виходу була того ж знака, що і реальний ряд. Цей позитивний результат показує, що вебтрафік вдається спрогнозувати за допомогою нейронної мережі. Однак спроби підвищити точність прогнозування на більш детальному рівні роздільної здатності вейвлету потребують додаткових досліджень.

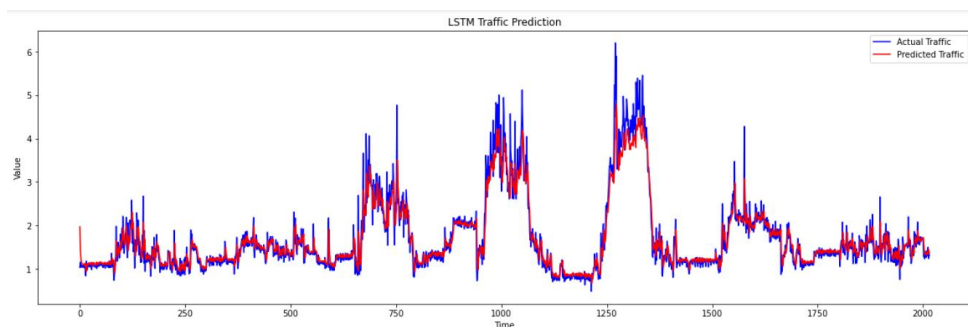


Рис. 4. Приклад досліджуваних тижневих коливань рівня вебтрафіка



## Висновки

Точний прогноз протягом певного часу вебтрафіку може значно допомогти у розробці програмного забезпечення, спрямованого на попередню вибірку, кешування та поширення вебдокументів [1]. Це також дозволило б нам генерувати синтетичні еталонні потоки з точними характеристиками продуктивності та очікувати зростання продуктивності для проектування систем. Якість доступних мережних підключень часто може мати великий вплив на продуктивність розподілених програм. Наприклад, програми передачі документів, такі як FTP і Web, потерпають від зростання часу відгуку в результаті перевантаження мережі. Для цих програм час передачі документа безпосередньо пов'язаний з доступною пропускну здатністю з'єднання. Якби оцінки потреби у трафіку були доступні за деякий час до запуску програми, можна було б розрахувати оцінку обсягу потрібних з'єднань. Потім наявні мережі можна було б використовувати як основу для вибору з набору альтернативних з'єднань або серверів, що забезпечило б скорочення часу відгуку. Ця можливість спонукала ряд авторів запропонувати надавати через додаток оцінку доступної пропускну спроможності між сервером та клієнтом, що дозволяє уникнути навантаження на рівні додатку [6].

Результати, представлені в цій статті, забезпечують прогнозування вебтрафіку з певною точністю на не дуже великих проміжках часу. Цей метод ще потребує серйозного доопрацювання, перш ніж можна буде систематично генерувати надійні пролонговані прогнози. В даний час проводиться подальше обґрунтування параметрів моделі за допомогою додаткових експериментів та аналізу даних вебтрафіку.

У цій статті ми досліджували прогнозованість вебтрафіку в різних масштабах часу за допомогою рекурентних нейронних мереж. Ми розглянули стратегію прогнозування, засновану на вейвлет-розкладі вихідного часового ряду на різні масштаби часового ряду, та застосували її до даних вебтрафіку. Вейвлет-перетворення

виконує вдалу декомпозицію даних, отже основні часові структури вихідного часового ряду стають зрозумілішими і чітко відображаються. Окремі вейвлет-коефіцієнти прогнозуються, потім рекомбінуються для отримання остаточного прогнозу. Показано, що усереднена кількість байтів, переданих за період одну годину, демонструє деяку передбачуваність під час використання цього методу.

## Література

1. Almeida V., Bestavros A., Crovella M. and de Oliveira A. Characterizing reference locality in the WWW / V. Almeida, A. Bestavros, M. Crovella and A. de Oliveira // DIS '96: Proceedings of the fourth international conference on Parallel and distributed information systems (December 1996) / – IEEE Computer Society, 1996. – P. 92-107.
2. Aussem A., Murtagh F. A neuro-wavelet strategy for Web traffic forecasting / A. Aussem, F. Murtagh // Research in Official Statistics. – 1998. – № 1. – P. 65-87.
3. Aussem A., Murtagh F. Web traffic demand forecasting using wavelet-based multiscale decomposition / A. Aussem, F. Murtagh // International Journal of Intelligent Systems. – 2001. – Vol. 16, Iss. 2. – P. 215-236.
4. Bruce, A. and Gao, H.-Y. S+Wavelets User's Manual, Version 1.0 / A. Bruce and H.-Y. Gao. – Seattle, WA:StatSci Division, MathSoft Inc., 1994.
5. Crovella M.E. and Bestavros A. Self-similarity in world wide web traffic evidence and possible causes / M.E. Crovella and A. Bestavros // SIGMETRICS '96: Proceedings of the 1996 ACM SIGMETRICS international conference on Measurement and modeling of computer systems (May 1996) / – Association for Computing Machinery, 1996. – P. 160-169.
6. Daubechies I. Ten Lectures on Wavelets / I. Daubechies. – Society for Industrial and Applied Mathematics, Philadelphia, 1992. – 343 p.
7. Holschneider M. and Tchamitchian P. Les ondelettes en 1989, ed. PG Lemarie / M. Holschneider and P. Tchamitchian. – Berlin: Springer-Verlag, 1990. – 102 p.
8. Leland W. and Wilson D.V. High time-resolution measurements and analysis of

LAN traffic: implications for LAN interconnections / W. Leland and D.V. Wilson. – IEEE Infocomm'91, 1991. – 12 p.

9. Lin T., Horne B.G., Tino P. and Giles C.L. Learning long-term dependencies in NARX recurrent neural networks / T. Lin, B.G. Horne, P. Tino and C.L. // IEEE Transactions on Neural Networks, 1996. – Vol. 7, № 6. – P. 1329-1338.

10. Starck J.-L. and Bijaoui A. Filtering and deconvolution by the wavelet transform / J.-L. Starck and A. Bijaoui // Signal Processing, 1994. – Vol. 35, Iss. 3. – P. 195-211.

11. Murtagh F., Starck J.-L. and Bijaoui A. Multiresolution in astronomical image processing: a general framework / F. Murtagh, J.-L. Starck and A. Bijaoui // International Journal of Imaging Systems and Technology, 1995. – Vol. 6, Iss. 4. – P. 332-338.

12. Starck J.-L., Murtagh F. and Bijaoui A. Multiresolution support applied to image filtering and deconvolution / J.-L. Starck, F. Murtagh and A. Bijaoui // Graphical Models and Image Processing, 1995. – Vol. 57, Iss. 5. – P. 420-431.

13. Wan E.A. Finite Impulse Response Neural Networks with Applications in Time Series Prediction: Ph.D. Thesis / E.A. Wan; Department of Electrical Engineering, Stanford University. – 1993. – 153 p.

14. Willinger W., Taqqu M., Leland W.E. and Wilson D. Self-similarity in high-speed packet traffic: Analysis and modeling of Ethernet traffic measurements / W. Willinger, M. Taqqu, W.E. Leland and D. Wilson // Statistical Science, 1995. – Vol. 10, № 1. – P. 67-85.

**Радченко К.О.**

## **ОСОБЛИВОСТІ ПРОГНОЗУВАННЯ РІВНЯ ВЕБТРАФІКУ У КОМП'ЮТЕРНИХ МЕРЕЖАХ ЗАГАЛЬНОГО ПРИЗНАЧЕННЯ**

*У статті проведено дослідження процесу прогнозування рівня вебтрафіку комп'ютерних мереж. Розглянуто можливість прогнозувати вебтрафік в різних масштабах часу за допомогою рекурентних нейронних мереж, використовуючи вейвлет-розклад вихідного часового ряду. Вейвлет-перетворення виконує декомпозицію даних, отже основні часові структури вихідного часового ряду стають чітко видимими. Окремі вейвлет-коефіцієнти прогнозуються, потім рекомбінуються для отримання остаточного прогнозу. Показано, що усереднена кількість байтів, переданих за період одну годину, демонструє прогнозованість під час використання цього методу. Проводиться подальше обґрунтування параметрів моделі за допомогою додаткових експериментів та аналізу даних вебтрафіку.*

**Ключові слова:** комп'ютерна мережа, рекурентна нейронна мережа, вейвлет, параметр Херста, часовий ряд, вебтрафік.

**Radchenko K.O.**

## **PECULIARITIES OF FORECASTING THE LEVEL OF WEB TRAFFIC IN GENERAL PURPOSE COMPUTER NETWORKS**

*The article studies the process of predicting the level of web traffic in computer networks. The possibility of predicting web traffic on different time scales using recurrent neural networks using the wavelet schedule of the original time series is considered. The wavelet transform decomposes the data so that the underlying temporal structures of the original time series become clearly visible. The individual wavelet coefficients are predicted, then recombined to obtain the final prediction. It is shown that the average number of bytes transferred in one hour demonstrates predictability when using this method. Further justification of the model parameters is carried out with the help of additional experiments and analysis of web traffic data.*

**Keywords:** computer network, recurrent neural network, wavelet, Hurst parameter, time series, web traffic.