

DOI: <https://doi.org/10.36910/6775-2524-0560-2021-45-13>

УДК 004.05

Радченко Костянтин Олександрович, асистент

<https://orcid.org/0000-0002-1282-6307>

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

ЗАСТОСУВАННЯ ДИСКРЕТНИХ ВЕЙВЛЕТ-ПЕРЕТВОРЕНЬ ДЛЯ ПРОГНОЗУВАННЯ РІВНЯ НАВАНТАЖЕННЯ НА ВЕБСЕРВЕР КОМП'ЮТЕРНИХ МЕРЕЖ ЗАГАЛЬНОГО ПРИЗНАЧЕННЯ

Радченко К. О. Застосування дискретних вейвлет-перетворень для прогнозування рівня навантаження на вебсервер комп'ютерних мереж загального призначення. Дана стаття присвячена застосування дискретних вейвлет-перетворень для розробки ефективного методу прогнозування навантаження, яке визначається розміром веб-трафіка, на вебсервер комп'ютерних мереж загального призначення. Проведено аналіз існуючих методів прогнозування у комп'ютерних мережах та запропоновано критерій для способу підвищення точності і швидкодії прогнозування навантаження. Проведено експеримент та визначено найбільш ефективний метод прогнозування навантаження на вебсервер.

Ключові слова: вейвлет-перетворення, вебсервер комп'ютерних мереж загального призначення, прогнозування навантаження, вейвлети Добеші.

Радченко К. О. Применение дискретных вейвлет-преобразований для прогнозирования уровня нагрузки на вебсервер компьютерных сетей общего назначения. Данная статья посвящена применению дискретных вейвлет-преобразований для разработки эффективного метода прогнозирования нагрузки, определяемого размером веб-трафика, на веб-сервер компьютерных сетей общего назначения. Проведен анализ существующих методов прогнозирования в компьютерных сетях и предложен критерий для способа повышения точности и быстродействия прогнозирования нагрузки. Проведен эксперимент и определен наиболее эффективный метод прогнозирования нагрузки на веб-сервер.

Ключевые слова: вейвлет-преобразование, веб-сервер компьютерных сетей общего назначения, прогнозирование нагрузки, вейвлеты Добеша.

Radchenko K. Application of discrete wavelet transforms for forecasting the level of load on the web server of general purpose computer networks. This article focuses on the application of discrete wavelet transforms to develop an effective method of predicting the load, which is determined by the amount of web traffic, on the web server of general-purpose computer networks. The analysis of existing forecasting methods in computer networks is carried out and the criterion for a way of increase of accuracy and speed of forecasting of loading is offered. An experiment was performed and the most effective method of forecasting the load on the web server was determined.

Keywords: wavelet transform, web server of general purpose computer networks, load forecasting, Daubechies wavelets.

Вступ. Задача прогнозування навантаження на вебсервер – складова частина системи захисту інформації, що визначає роботу вебсервера, який обслуговує запити від віддалених клієнтів браузера з метою забезпечення безперебійного доступу джерел інформації.

Зауважимо, що найпопулярніший вебсервер Apache за стандартною конфігурацією не може одночасно обслуговувати більше 30 клієнтів. Задача прогнозування навантаження на вебсервер може бути вирішена шляхом розробки відповідних моделей. У зв'язку з тим, що процес підключення віддалених клієнтів до комп'ютерної мережі має чітко виражений випадковий характер, то доцільно використовувати статистичні моделі для прогнозування навантаження.

Очевидно, що тип і параметри моделі прогнозу в значній мірі залежать від тих параметрів, які визначають навантаження на вебсервер. Базуючись на типових завданнях, що виконуються вебсервером, а також на протоколах передачі даних, навантаження можна оцінити за групами параметрів, які характеризують використання:

1. Апаратні ресурси комп'ютер-сервер – оперативна пам'ять, жорсткий диск, мікропроцесор.

2. Ресурси операційної системи – довжина черги мережі, кількість відкритих файлів, кількість виконуваних процесів тощо.

3. Мережеві ресурси – обсяг вхідного та вихідного трафіку, розмір мережевих пакетів, кількість клієнтів на одиницю часу, кількість "невідповідних" мережевих пакетів для кожного з використовуваних мережевих протоколів тощо.

Одним з найсучасніших методів частотно-часового аналізу нестационарних процесів є вейвлет-перетворення, яке дозволяє розрахувати частотно-часові характеристики вебсерверів. Однак ця модель заснована на теорії неперервного вейвлет-перетворення, і тому має значну надмірність і вимагає адаптації до дискретних даних, які відповідають типовій статистиці параметрів вебсервера.

Можливим ефективним способом усунення цих недоліків може бути використання дискретного вейвлет-перетворення.

Пов'язані роботи. Як правило у ранніх роботах [2], [4], [5], [8] для навантаження на вебсервер у часі використовувалися моделі прогнозування задані простими аналітичними виразами, а саме лінійна, поліноміальна, експоненціальна, ковзного середнього, експоненціального згладжування тощо. Їх основною перевагою є простота математичного апарату. У той же час практичний досвід їх використання показує серйозні недоліки. У разі реалізації задачі прогнозування навантаження ці моделі не дозволяють адекватно враховувати нестационарний і багатоперіодичний характер типових залежностей більшості експлуатаційних параметрів вебсервера від часу.

У роботах [12], [13] з метою підвищення точності предикатів веб-трафіку, був запропонований метод предикації на основі інтеграції вейвлет-аналізу та нейронних мереж. Часовий ряд веб-трафіку, який є нелінійним і нестационарним, був розкладений і потім перебудований у кілька гілок методом вейвлета. Ці гілки були передбачені нейронними мережами причому кінцеве значення було комбінацією цих прогнозованих результатів.

Проведений теоретичний аналіз та результати експериментів показують, що вейвлет-аналіз може розкласти вихідний ряд трафіку на кілька часових рядів, які мають простіші частотні компоненти і легші у прогнозуванні. Таким чином, показано, що розроблений метод має більш високу предиктивну точність, ніж традиційні підходи прогнозування. У роботі [16] представлена модель прогнозу навантаження на вебсервер з урахуванням аналізу вейвлет-пакетів. Модель припускає, що часові ряди динамічного рівня веб-трафіка для вебсервера розкладаються та реконструюються за допомогою аналізу вейвлет-пакетів, щоб отримати набір вейвлет-коефіцієнтів однакової довжини за певний період часу. Експериментальним шляхом доводиться, що запропонований метод перевершує традиційний підхід до прогнозування.

У роботі [6] розроблено підхід для прогнозування короткострокового навантаження. Підхід використовує векторні машини (LS-SVM) та теорію вейвлет-перетворень. Часовий ряд аналізується за допомогою вейвлет-перетворень, результати якого прогноуються окремою модифікацією LS-SVM. Нова модель прогнозу поєднує перевагу вейвлет-перетворень з LS-SVM. Задекларовано, що в порівнянні з іншими предикторами ця модель прогнозу має більшу узагальнюючу здатність і більш високу точність.

В результаті проведеного аналізу [17], [18]], [19] можна стверджувати, що застосування вейвлет-перетворень є одним із найбільш перспективних напрямків підвищення ефективності систем прогнозування навантаження на вебсервер. При цьому теорія вейвлет-перетворень свідчить, що неправильний вибір цих параметрів може в значній мірі знизити ефективність системи прогнозування.

З точки зору дискретного вейвлет-аналізу, статистичні дані вебсервера мають такі особливості: обмежені інтервалом $t \in [t_{min}, t_{max}]$, дані реєструються з певною дискретністю Δt , для діадного перетворення кількість точок ряду N має дорівнювати $N = 2z$ (z – ціле число).

Модель прогнозування рівня веб-трафіку вебсервера комп'ютерних мереж загального призначення. Представлена модель базується на принципі визначення очікуваного рівня веб-трафіка вебсервера комп'ютерних мереж загального призначення, процедурах експертного прогнозування рівня веб-трафіка та його масштабування для досліджуваного вебсервера. Основою математичного забезпечення моделі є вирази для прямого та зворотнього вейвлет-перетворення:

$$W(m, k) = 2^{0.5m} \sum_{i=1}^N (f(t_i) \psi^*(2^m t_i - k)),$$

$$f(t_i) = \frac{\pi}{\ln 2} \sum_{m=0}^{N-1} \sum_{k=0}^{L-1} \psi(t) W(m, k),$$

де N – кількість точок ряду, L – кількість масштабів, m – параметр масштабу, k – параметр зсуву, ψ – базисний вейвлет, $*$ – процедура комплексного спряження, $f(t_i)$ – відновлення дискретної функції по набору вейвлет-коефіцієнтів у момент t_i .

Задача прогнозування формулюється наступним чином: маючи наявні попередні l значень часового ряду $x(t)$, передбачити наступні значення $x(t+d)$:

$$x(t+d) = \Psi(x(t), (t-1), \dots, (t-l), \xi_1, \xi_2, \dots, \xi_q),$$

де d – крок прогнозування; q – кількість незалежних змінних ξ ; Ψ – вейвлет-функція, що використовується для прогнозування.

Архітектура такої моделі безпосередньо залежить від множини тих параметрів, які власне і визначають навантаження на вебсервер. Виходячи з типових задач, які виконуються вебсервером комп'ютерних мереж загального призначення, а також використовуваних мережевих комунікаційних

протоколах визначені три групи параметрів, кожна з яких характеризують використання: апаратних ресурсів комп'ютера-сервера; ресурсів операційної системи; мережевих ресурсів.

Для визначення довгострокового прогнозу треба проаналізувати природу подій, їх можливий зв'язок з рядом інших подій, легко прогнозованих.

Зокрема, аналізуючи роботу вебсервера комп'ютерних мереж загального призначення потрібно врахувати зв'язок з ритмами життя людей, які користуються досліджуванним вебсервером. Має простежуватись добова періодичність, тижнева, сезонна, річна. Таким чином потрібно мати дані спостережень за період, який становить не менше 1 року. Виявлення періодів та їх повторюваність – це вже ключ до прогнозу. Оскільки спостереження відбуваються у часі, то і отриманий трафік навантаження на вебсервер також будемо розглядати як функцію, що залежить від часу.

$$\langle S, IO, D, K, K_t, K_s, T, Z_{nb}, W_n \rangle \rightarrow \langle Z \rangle,$$

$$Z = \{z^{(\phi_k)}\}_{K_\phi} = \left\{ \left(x^{(\phi_k)}, y^{(\phi_k)} \right) \right\}_{K_\phi},$$

де S – множина спостережуваних типів сервера, K – множина зареєстрованих клієнтів, IO – множина співвідношення вхідного/вихідного трафіку в залежності від типу дня (робочий, вихідний, передсвятковий, святковий, рухоме свято тощо), D – множина несподіваних збурень вебтрафіку даних, згідно зібраної статистики, K_t – множина середньої тривалості користування вебсервером серед клієнтів, K_s – множина середньої кількості перегляданих вебсторінок, T – тренд зміни навантаження на вебсервер за період спостереження, Z – статистична вибірка, приклади якої містять періодичні (доба, тиждень, рік) особливості досліджуваного вебтрафіку, Z_{nb} – множина обмежень на формування прогнозованої вибірки, W_n – обчислювальна потужність вебсерверу, $z^{(\phi_k)}$ – пронормовані приклади виду Z для k -го досліджуваного елемента, $x^{(\phi_k)}$ – множина вхідних параметрів для k -го елемента, $y^{(\phi_k)}$ – очікуваний вихідний сигнал для k -го елемента, K_ϕ – кількість елементів, які мають бути враховані.

Вибір факторів для кожного випадку є окремою задачею, що вимагає заздалегідь оцінити вплив прогнозованих факторів прогнозування до початку самого прогнозування.

Якість прогнозування оцінювалася показником середньо-відсоткової похибки MAPE (Mean Absolute Percentage Error):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{Z}_t - Z_t|}{Z_t} 100\%,$$

де \hat{Z}_t – прогнозоване значення, Z_t – фактичний показник веб-трафіку, n – розмір вибірки.[2]

Практичний досвід та зібрані результати вказують на те, що ефективність більшості вітчизняних корпоративних інформаційних систем безпосередньо залежить від якості функціонування вебсерверу, за допомогою якого забезпечуються відповіді на запити віддалених клієнтів. Очевидно, що обчислювальні ресурси вебсерверу повинні відповідати його навантаженості. В протилежному випадку вебсервер або не зможе адекватно забезпечувати потреби користувачів інформаційної системи, або резервування надлишкових потужностей призведе до економічних та енергетичних втрат.

Експериментальна оцінка. В ході дослідження були проведені порівняльні оціночні експерименти наведених вище методів прогнозування навантаженості на вебсервер з використанням статистики досліджуваного вебсервера за звітний період з липня 2019 року по червень 2021 року.

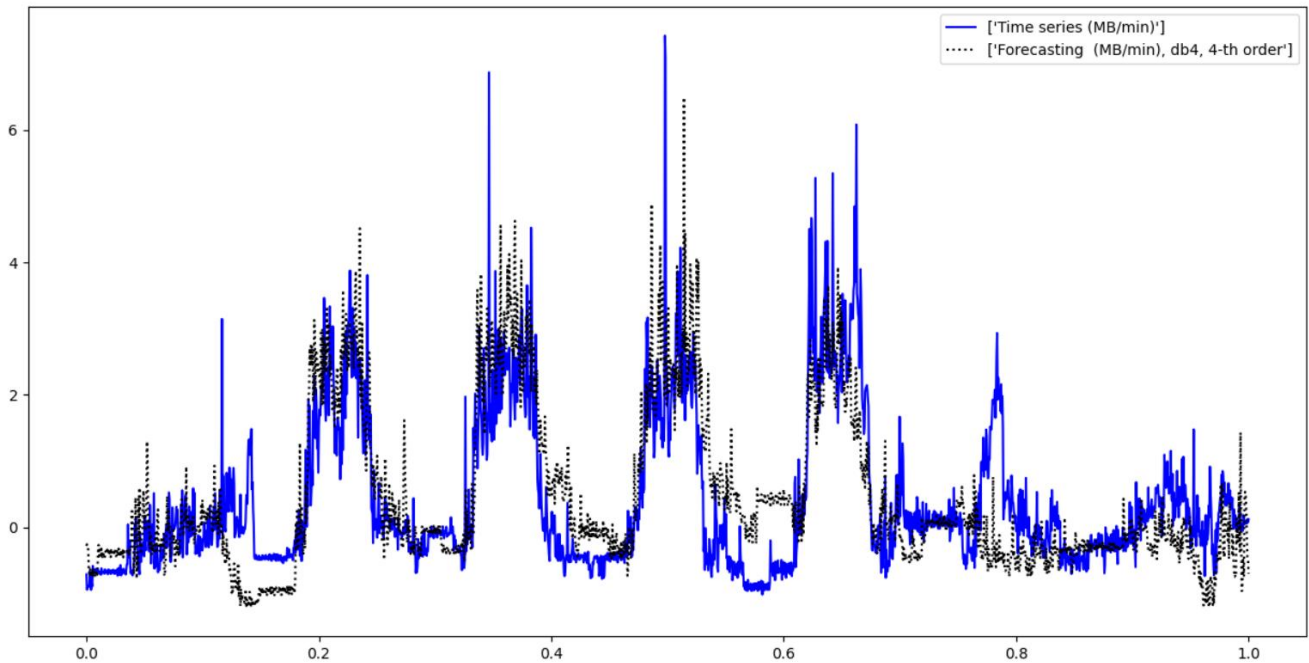


Рис.1. Експериментальні дослідження методу застосування вейвлет-перетворень для прогнозування короткострокових (тижневих) навантажень на вебсервер

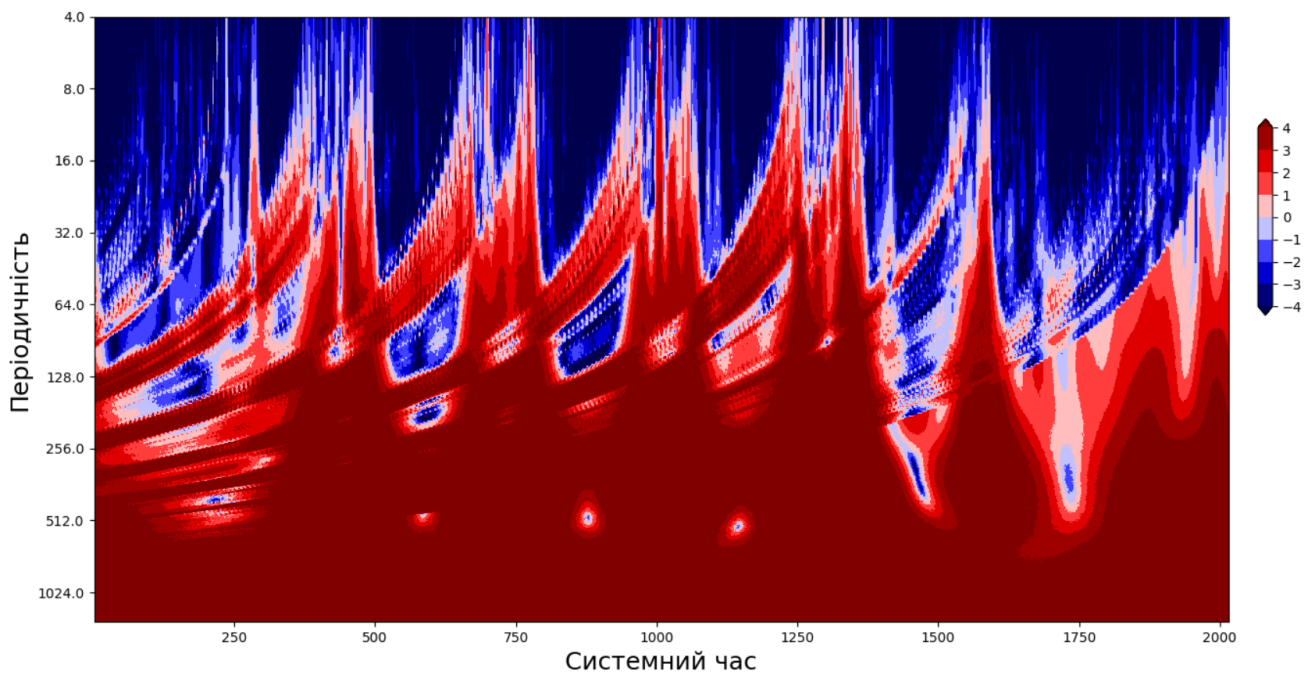


Рис.2. Скалограма вейвлет-перетворень для прогнозування короткострокових (тижневих) навантажень на вебсервер

На реальних даних було проведено вейвлет-аналіз Добеші статистики відвідуваності вебсайту. Розрахунки проводились за допомогою бібліотеки PyWavelets мови програмування Python. Аналіз даних вказує на те, що в досліджуваному процесі існують періодичні компоненти, які дорівнюють добовим, тижневим і сезонним коливанням.

Перед початком експерименту було проведено нормування даних. Дані були завантажені, розбиті на окремі періоди, очищені від аномальних викидів веб-трафіку.

Прогнозування навантаження проводилося у різних часових діапазонах. У роботі часові діапазони прогнозування класифікувалися наступним чином: у межах поточної доби – оперативний прогноз; від 1 доби до тижня наперед – короткостроковий прогноз; від тижня до 1 місяця наперед – середньостроковий прогноз; довгостроковий прогноз – більше 1 місяця до кварталу і більше.

На Рис. 1. можна побачити результати застосування методу дискретних вейвлет-перетворень для прогнозування короткострокових (тижневих) навантажень на досліджуваний вебсервер, а на Рис. 2 можна побачити представлення цих же вейвлет-перетворень у вигляді скалограми (трьохвимірне представлення значень одержаних вейвлет-коефіцієнтів). При аналізі екстрапольованого часового ряду середньо-відсоткова помилка апроксимації становила (2 – 3)%.

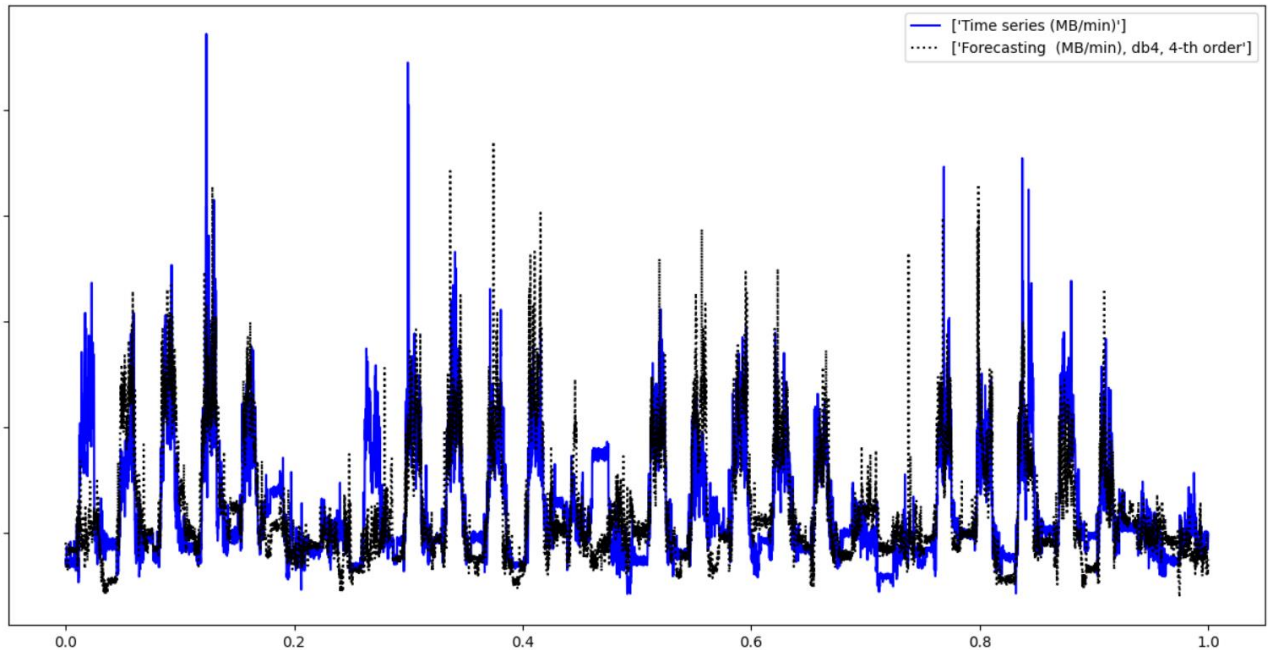


Рис.3. Експериментальні дослідження методу застосування вейвлет-перетворень для прогнозування середньострокових (місячних) навантажень на вебсервер

На Рис. 3. можна побачити результати застосування методу дискретних вейвлет-перетворень для прогнозування середньострокових (місячних) навантажень на досліджуваний вебсервер, а на Рис. 4 представлений той же досліджуваний період без урахування аномальних викидів трафіку, причому середньо-відсоткова помилка апроксимації становила (3 – 5)%.

На Рис. 5. можна побачити результати застосування методу дискретних вейвлет-перетворень для прогнозування довгострокових (квартальних) навантажень на досліджуваний вебсервер, причому середньо-відсоткова помилка апроксимації становила (5 – 7)%.

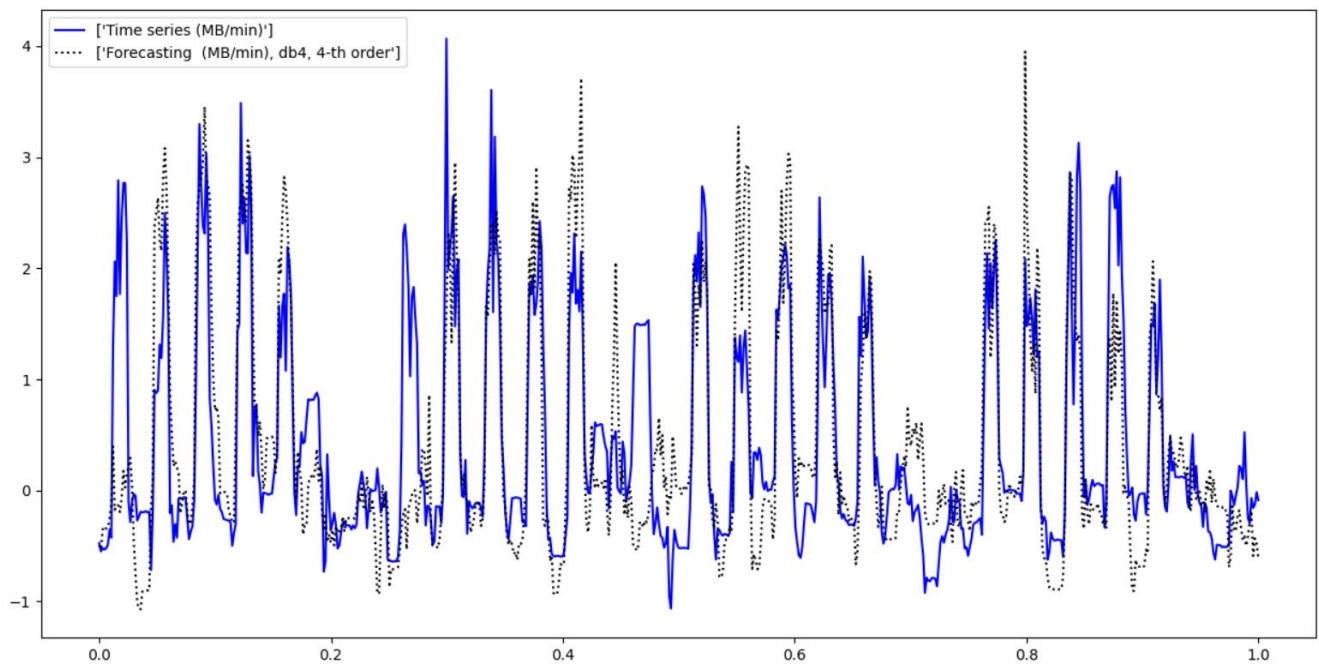


Рис. 4. Експериментальні дослідження методу застосування вейвлет-перетворень для прогнозування середньострокових (місячних) навантажень на вебсервер, без урахування аномальних викидів трафіку

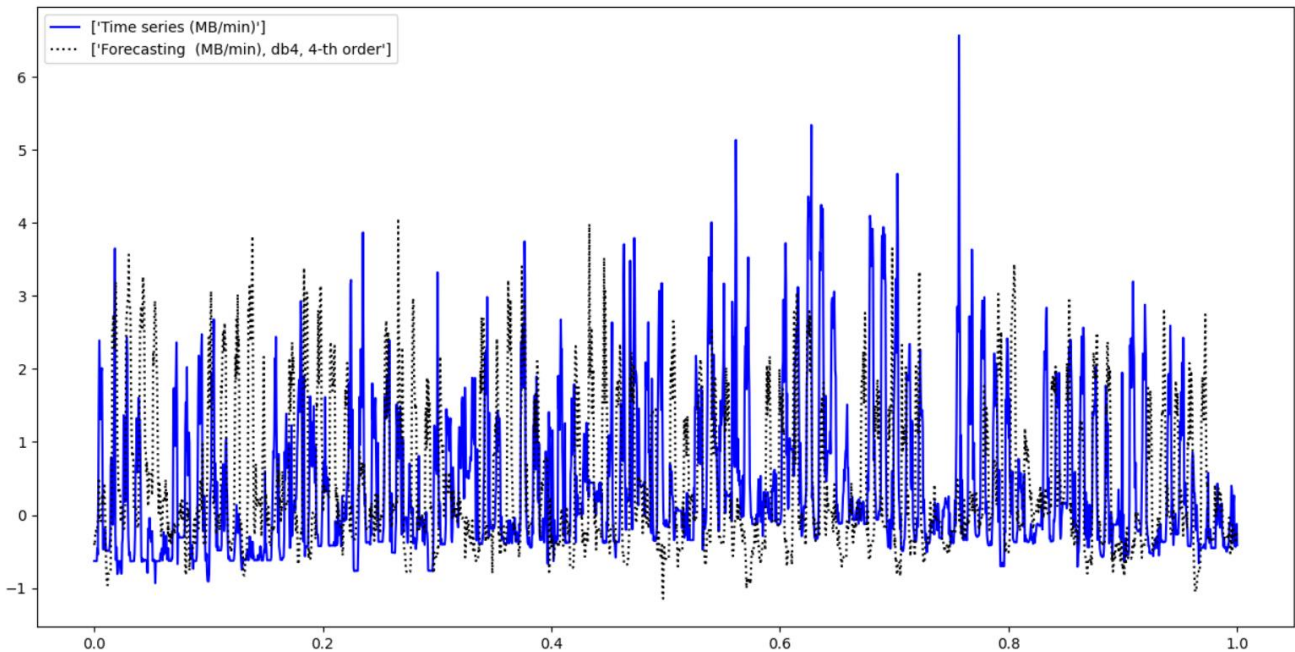


Рис. 5. Експериментальні дослідження методу застосування вейвлет-перетворень для прогнозування довгострокових (квартальних) навантажень на вебсервер

Для числових експериментів було застосовано різні набори ортогональних вейвлетів сімейства Добеші, що дозволило верифікувати отримані результати шляхом перевірки точності апроксимації статистичних даних та можливості використовувати вейвлет-коефіцієнти для визначення локальних особливостей динаміки рівня навантаження вебсервера:

$$\psi(t) = \sqrt{2} \sum_{k=1}^Q h_k \psi^*(2t - k),$$

де h_k – ортогональні нормовані коефіцієнти Добеші. Як найбільш ефективний тип базисного вейвлету цього сімейства було обрано Daubechies-5 з коефіцієнтами:

$$\begin{cases} h_1 = 0.0033357253; \\ h_2 = -0.0125807520; \\ h_3 = -0.0062414902; \\ h_4 = 0.0775714938; \\ h_5 = -0.0322448696; \end{cases} \begin{cases} h_6 = -0.2422948871; \\ h_7 = 0.1384281459; \\ h_8 = 0.7243085284; \\ h_9 = 0.6038292698; \\ h_{10} = 0.1601023980. \end{cases}$$

Отже, отриманий набір вейвлет-коефіцієнтів дозволяє екстраполювати періодичні особливості досліджуваного часового ряду веб-трафіка на період часу, що перевищує 1 квартал, причому при аналізі одержаного прогнозованого часового ряду середньо-відсоткова помилка апроксимації не перевищує (7 – 10)%.

Висновок

В даній статті були описані методи застосування дискретних вейвлет-перетворень для розробки ефективного методу прогнозування навантаження, яке визначається розміром веб-трафіка, на вебсервер комп'ютерних мереж загального призначення. Одержаний результат дає кращі результати порівняно з класичними поліноміальними методами прогнозування, менш ресурсоємний та дозволяє виконувати прогноз на більшу кількість кроків порівняно з рекурентними нейронними мережами, зокрема LSTM, що дозволяє успішно використовувати його для прогнозування нестационарного навантаження на вебсервер.

Таким чином результати експериментів підтверджують можливість застосування розробленого методу, перспективи вдосконалення якого полягають в уточненні розрахунку критеріїв ефективності і їх вагових коефіцієнтів. Крім того представляє інтерес деталізація запропонованого методу в напрямку розрахунку параметрів базисного вейвлета. Запропоновані моделі вимагають додаткового дослідження та можуть стати основою для подальших робіт.

Список бібліографічного опису.

1. Довлад О. А. Дослідження та розробка моделі процесу атаки та трафіку локальної мережі / О. А. Довлад // *Захист інформації*. – 2009. – № 1 – Р. 83 – 86.
2. Менаске Д. Производительность Web-служб. Анализ, оценка и планирование / Менаске Д., Виргилио А. ; пер. с англ. – СПб.: "ДиаСофтЮп", 2003, 480 с.
3. Переберин А. В., О классификации вейвлет-преобразований, Выч. мет. программирование, 2:3 (2001), 15–40.
4. Сыропятов А. А. Метод мониторинга трафика защищенных высокоскоростных коммерческих сетей нового поколения / А. А. Сыропятов // *Наукові записки УНДІЗ*. – 2009. – № 2(1) – С. 65-73.
5. Терейковская Л. А. Применение дискретного вейвлет-преобразования для определения частотно-временных параметров эксплуатационной нагрузки Web-сервера системы дистанционного образования [Текст] // *Науково-технічний збірник «Управління розвитком складних систем» КНУБА*. – 2011. – Вип. 5. – С. 124 – 128.
6. Дичка І., Терейковський І., Терейковська Л., Радченко К. Метод визначення ефективного типу базисного вейвлету для застосування в шаблонах нормальної поведінки веб-сервера // *Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні*. – № 2 (36). – 2018. – С. 46 – 55.
7. Радченко К.О. Концептуальна модель забезпечення ефективності прогнозування навантаження на вебсервер // *Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки*. – Том 31 (70). – № 6. – 2020. – С. 135 – 141.

References

1. Bariyev, I.M., Aitchanov, B.H., Tereikovskiy, I.A., Tereikovska, L.A., Korchenko, A.A. (2017), "Deep neural networks in cyber attack detection systems", *International Journal of Civil Engineering and Technology (IJCIET)*, Volume 8, Issue 11, November 2017, pp. 1086–1092.
2. Dychka, I., Tereikovskiy, I., Tereikovska, L., Pogorelov, V., Mussiraliyeva, S.(2018), "Deobfuscation of computer virus malware code with value state dependence graph", *Advances in Intelligent Systems and Computing*, pp 370-379, 2018. DOI: 10.1007/978-3-319-91008-6.
3. Qi-Song Chen ; Xin Zhang ; Shi-Huan Xiong ; Xiao-Wei Chen Short-term power load forecasting with least squares support vector machines and wavelet transform 2008 International Conference on Machine Learning and Cybernetics Year: 2008 , Volume: 3. Pages: 1425 – 1429.
4. Sitnikov V.S., Bilenko A.A.. Classification of wavelet functions. Труды Одесского политехнического университета, 2008, вып. 1(29) с.168-171.
5. Starck, J.-L. Image and Data Analysis: the Multiscale Approach / J.-L. Starck, F. Murtagh, A. Bijaoui. — Cambridge University Press, 1998. — 307 с.
6. Steinbuch, M. Wavelet Theory and Applications A literature study / M. Steinbuch, M.J.G. van de Molengraft — Eindhoven University of Technology, 2005. - 39 с
7. Shuping Yao; Changzhen Hu; Mingqian Sun Prediction of Web Traffic Based on Wavelet and Neural Network 2006 6th World Congress on Intelligent Control and Automation Year: 2006 , Volume: 1, Page s: 4026 – 4028.
8. Shuping Yao; Changzhen Hu; Wu Peng Server Load Prediction Based on Wavelet Packet and Support Vector Regression 2006 International Conference on Computational Intelligence and Security Year: 2006, Volume: 2 Pages: 1016 – 1019.
9. Tereykovska, L., Tereykovskiy, I., Aytkhozhaeva, E., Tynymbayev, S., Imanbayev, A. (2017), "Encoding of neural network model exit signal, that is devoted for distinction of graphical images in biometric authenticate systems", *News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences*, Volume 6, Number 426, pp. 217 – 224, 2017.
10. Zhengbing Hu, Igor A. Tereykovskiy, Lyudmila O. Tereykovska, Volodymyr V. Pogorelov, "Determination of Structural Parameters of Multilayer Perceptron Designed to Estimate Parameters of Technical Systems", *International Journal of Intelligent Systems and Applications(IJISA)*, Vol.9, No.10, pp.57-62, 2017. DOI: 10.5815/ijisa.2017.10.07.
11. Zijiang Yang Research on Server Load Prediction Based on Wavelet Packet Theory 2007 First IEEE International Symposium on Information Technologies and Applications in Education Year: 2007 Pages: 610 – 613.
12. Hu Z., Tereykovskiy I., Tereykovska L., Tsiutsiura M., Radchenko K. Applying Wavelet Transforms for Web Server Load Forecasting // *Advances in Computer Science for Engineering and Education II. ICCSEEA 2019. Advances in Intelligent Systems and Computing*, vol. 938, pp. 13–22 (2019). Springer, Cham. ISSN 2194-5365 (Online). https://doi.org/10.1007/978-3-030-16621-2_2