

Високопродуктивний алгоритм зіставлення патернів у часових рядах для задач аналізу якості потокових сервісів

Дорошенко Анатолій Юхимович
доктор фізико-математичних наук, професор
КПІ ім. Ігоря Сікорського
Україна, Київ

Тітов Дмитро Сергійович
аспірант
КПІ ім. Ігоря Сікорського
Україна, Київ

Розглянуте застосування алгоритмів та технік аналізу часових рядів в задачах моніторингу та прогнозування якості передачі даних в потокових сервісах, таких як VoIP. Запропоноване покращення алгоритму знаходження подібних патернів часових послідовностей з точки зору продуктивності обчислень.

Ключові слова: VoIP; QoS; MOS; джитер; патерн; часовий ряд; PAA; SAX.

1. ВСТУП

Практична проблема забезпечення якості обслуговування (Quality of Service, QoS) істотних обсягів різнорідного мережевого трафіку в умовах невизначеності динаміки мережевих потоків та мережевого середовища набуває важливого значення у сучасних телекомунікаційних мережах [1]. QoS є необхідним для пакетних мереж, які використовуються для сервісів працюючих у режимі реального часу, насамперед VoIP. Мережеві протоколи, які обслуговують сервіси реального часу є чутливими до якості обслуговування, а саме до втрати пакетів даних, затримок у передачі пакетів та нерівномірності цих затримок [2].

Для оцінювання якості надання послуг використовують показник оцінювання якості сприйняття послуг QoE (Quality of Experience), який прямопропорційно залежить від показника якості надання сервісу QoS. В свою чергу, для оцінки QoE найчастіше використовують величину MOS (Mean opinion score). MOS – це середнє арифметичне по всіх окремих «значеннях за задалегідь визначеною шкалою, що суб'єкт приписує його враженню про продуктивність системи в цілому». Такі рейтинги, як правило, зібрані в тесті суб'єктивної оцінки якості, але вони також можуть бути алгоритмічно оцінені [3].

2. ОГЛЯД РОБИТ

Існує дуже багато праць, присвячених різноманітним підходам до аналізу трафіку, алгоритмам їх моніторингу та навіть передбачення показників якості. Наприклад, роботи [4-7] демонструють, що VoIP-трафік має властивості самоподібності та може містити у собі довготривалі залежності, а у роботі [8] розглядається побудова QoS моделей VoIP-трафіку, заснованих на мультифрактальності та Марковських моделях.

У роботі [9] розглянута можливість прогнозування продуктивності мережевих інфраструктур за допомогою Марковських процесів, а в [10] пропонується використання нейромережевих підходів та лінійної регресії, як технік прогнозування показників навантаження в мультисервісних мережах.

Низка робіт присвячена можливостям представлення VoIP-показників за допомогою нечіткої логіки [11] або передбаченню рівня QoS з використанням штучних нейронних мереж [12]. Також існують праці, що описують підходи до моніторингу QoE високорівневими засобами, наприклад, аналіз соціальних мереж у реальному часі [13].

3. ПАРАМЕТРИ VOIP-ТРАФІКУ, ЩО ВПЛИВАЮТЬ НА ЯКІСТЬ ПОСЛУГ

Голосові кодеки вимагають постійного, надійного потоку пакетів, щоб забезпечити прийнятну якість відтворення. Пакети, які прибувають занадто рано, занадто пізно, або з ривками, спотворюють відтворення аудіосигналу. Це явище називається джитером.

Джитер – це небажані фазові та/або частотні випадкові спотворення при передачі сигналу. У той час як джитер-буфер може успішно замаскувати невеликі проблеми затримки, суттєвий джитер може призвести до переповнення джитер-буферу, що в свою чергу призводить до втрати пакетів.

Найбільш важливим фактором для якості зв'язку є “burstiness” – так звані сплески. Сплески є періодами, що характеризуються високим рівнем втрати пакетів [14] (рис. 1).

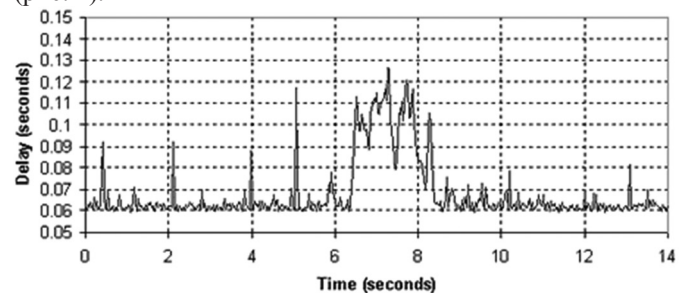


Рис. 1. Приклад спалаху на графіку затримки пакетів

4. ІДЕНТИФІКАЦІЯ ПАТЕРНІВ ПАРАМЕТРІВ ЯКОСТІ МЕРЕЖЕВОГО ТРАФІКУ

Дуже важливу роль в аналізі часових рядів, в тому числі в задачах моніторингу якості потокових сервісів, віді-

грає pattern matching – знаходження “подібних” фрагментів у послідовності даних (рис. 2).

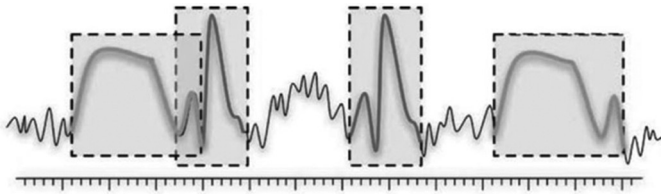


Рис. 2. Знаходження подібних фрагментів у часовому ряді

В цій роботі розглядається символічний алгоритм кодування часових послідовностей – SAX (Symbolic Aggregate approXimation). Він має низку переваг, таких як можливість хешування даних, застосування Марковських моделей, суфіксних дерев, дерев прийняття рішень та ін. [15]

Алгоритм SAX дозволяє описати часовий ряд довільної довжини n рядком довільної довжини w ($w < n$, як правило $w \ll n$). Рядок складається з символів певного алфавіту, розмір якого – довільне ціле число α , де $\alpha > 2$ [16].

Алгоритм відрізняється від інших [17] процедурою дискретизації, в якій використовується проміжне представлення часового ряду (РАА, Piecewise Aggregate Approximation).

Часовий ряд C довжини n може бути представлений у просторі розмірністю w вектором $\bar{C} = \bar{c}_1, \dots, \bar{c}_w$. i -й елемент \bar{C} обчислюється за наступною формулою:

$$\bar{c}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} c_j$$

Далі РАА-перетворенню ставиться у відповідність символна послідовність з використанням алгоритму SAX, який на основі квантилів наявного розподілу ставить у відповідність до кожного амплітудного рівня символ із попередньо визначеного алфавіту A . Квантилі відсікають в межах ряду певну частину його членів. Тобто, квантиль розподілення значень — це таке число x_p , що значення p -ї частини сукупності менше або рівне x_p (рис. 3).

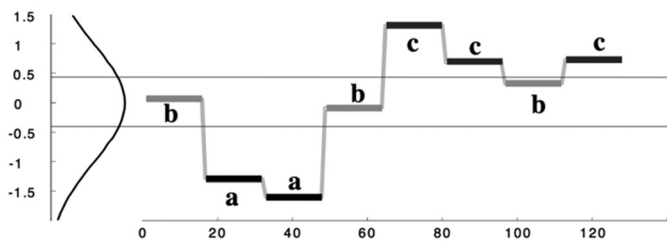


Рис. 3. Часовий ряд перетворюється у символну послідовність “baabcabc” з параметрами $n = 128$, $w = 8$, $\alpha = 3$

Далі для знаходження певних сталих патернів часовий ряд сканується відрізком певної довжини з заданим зсувом та кожний відрізок порівнюється з оригіналом (sliding window).

5. ПОКРАЩЕННЯ ПРОДУКТИВНОСТІ ЗНАХОДЖЕННЯ ПОДІБНИХ ПАТЕРНІВ У ЧАСОВИХ РЯДАХ

Підхід “sliding window” має суттєвий недолік з огляду на продуктивність обчислень. Якщо стоїть задача для двох закодованих часових послідовностей знайти всі співпадаючі підпослідовності, доводиться виконувати повне сканування вихідних послідовностей декілька разів, оскільки попередньо ми не знаємо якої довжини подібні патерни можемо зустріти.

Очевидно, що в задачах реального часу, таких як аналіз VoIP-трафіку, така ітеративність алгоритму є непринятною, адже повний перебір усіх можливих вікон для співставлення даних унеможливить своєчасну обробку інформації.

Протягом досліджень нами було помічено, що якщо відобразити значення відстаней між усіма вікнами одного розміру у вигляді двовірної карти відстаней, у такому відображенні можна виділити геометричні примітиви, які вказують на подібність підпослідовностей значно більшої довжини, ніж довжина вікна, яка використовувалась для складання візуальної карти даної ітерації (рис. 4).

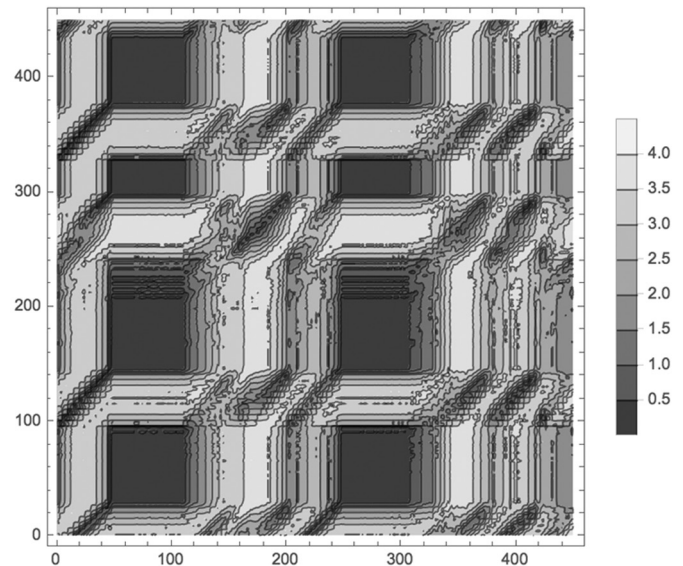


Рис. 4. Двовимірні карти відстаней підпослідовностей з розміром вікна 50

Якщо зіставити, використовуючи осі абсцис та ординат, відрізки вихідних послідовностей, які відповідають діагональним відрізкам малих відстаней, то можна побачити, що ці фрагменти послідовностей є подібними. Наприклад, на рисунку 4 відрізок між точками (20; 0) та (120; 100) відповідає наведеним нижче інтервалам аналізованих часових рядів (рис. 5).

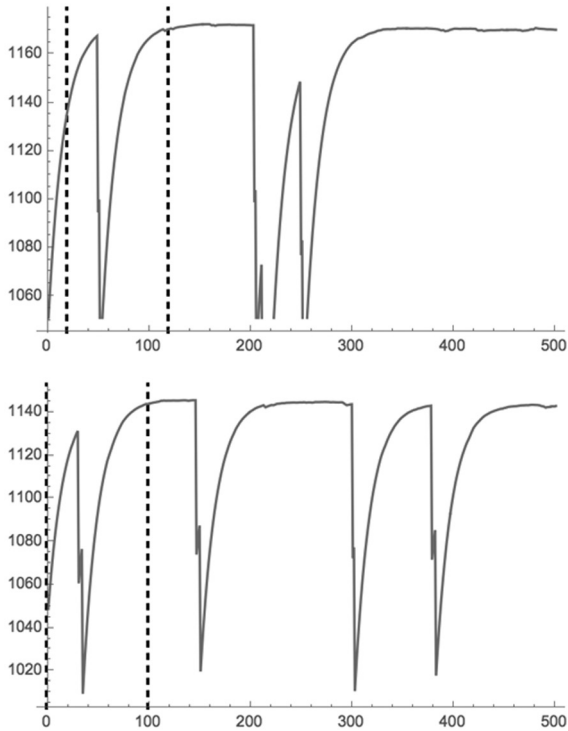


Рис. 5. Знайдені подібні підпоследовності у двох часових последовностях

Як бачимо, з використанням вікна однієї довжини (50) вдалося знайти подібні патерни довжиною 100 елементів (без необхідності у додаткових 50 ітераціях, щоб збільшити розмір вікна до 100).

6. ЕКСПЕРИМЕНТАЛЬНЕ ПОРІВНЯННЯ ПРОДУКТИВНОСТІ

КЛАСИЧНОГО ТА УДОСКОНАЛЕНОГО ПІДХОДІВ

Розроблені в процесі дослідження Java-додатки використовують спільні алгоритми для проведення РАА-та SAX-перетворень, проте різні процедури порівняння отриманих даних. Перший додаток для знаходження схожих часових патернів використовує почергове сканування даних вікнами змінної довжини (з 50 до 140), а другий базується на підході, запропонованому у п'ятому розділі. Тестові дані для аналізу взяті з реальної VoIP-сесії та екстрактовані за допомогою програми Wireshark (рис. 6).

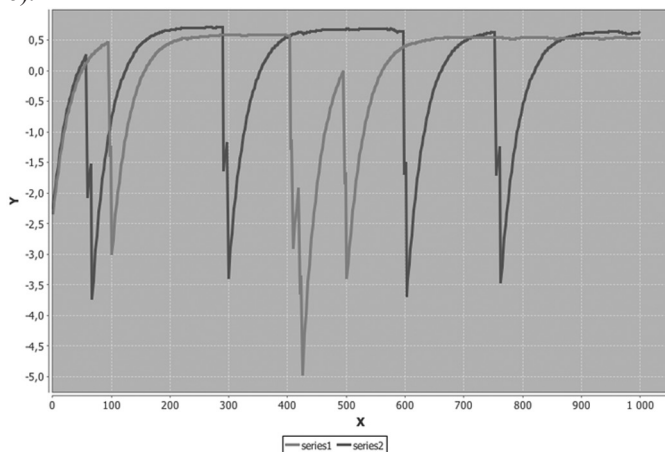


Рис. 6. Тестові дані – джитер

Обидві програми знайшли кілька співпадаючих патернів, найдовшим з яких є наступний (рис. 7).

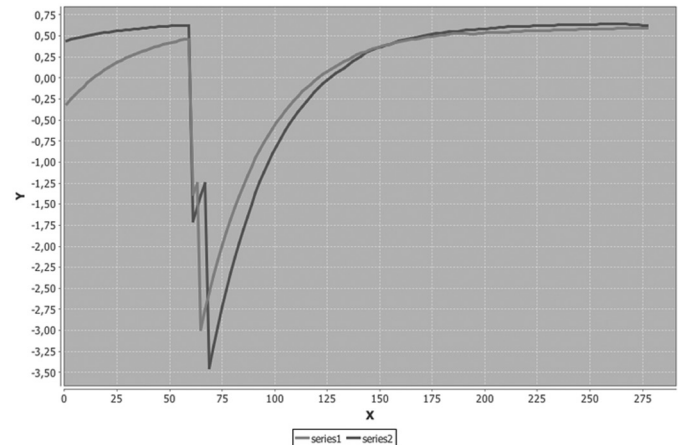


Рис. 7. Приклад знайденого патерну

Також ми провели тест існуючої реалізації SAX, що називається jMotif. Нижче наведений графік, який демонструє, скільки часу в мілісекундах пішло на сканування кожним новим вікном у класичному підході (власна імплементація та jMotif, 90 ітерацій), та скільки часу зайняв модернізований підхід (одна ітерація) (рис. 8).

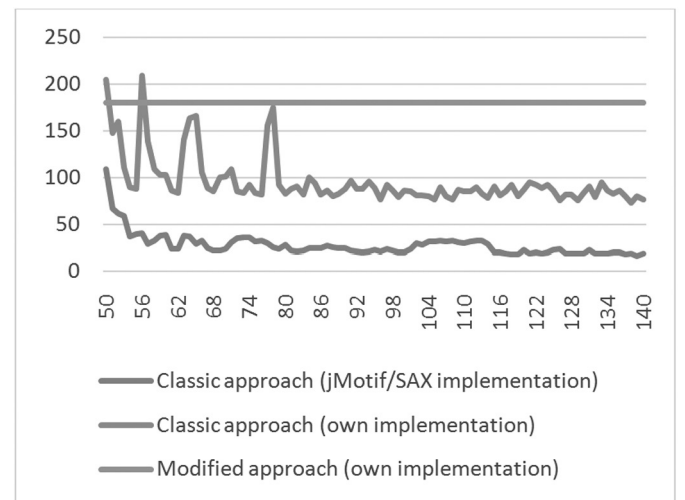


Рис. 8. Порівняльна діаграма швидкодії

Загалом повний цикл роботи програми, заснованої на класичному підході, зайняв 8740 мс та 2543 мс (власна та jMotif імплементація відповідно), у той час як виконання модернізованого алгоритму зайняло лише 180 мс, що вказує на приріст швидкодії більш ніж у 14 разів порівняно з jMotif та 48 разів порівняно з власною імплементацією.

ВИСНОВКИ

У роботі проаналізоване застосування алгоритмів та технік обробки часових рядів в задачах моніторингу та прогнозування якості передачі даних в потокових сервісах, таких як VoIP. Показано, які саме параметри впливають на якість аудіосигналу в голосових сесіях, та як можна їх представити у вигляді часового ряду.

Розглянуто символічний алгоритм кодування часових послідовностей SAX, який має низку переваг перед аналогами та оперує власною мірою відстаней між послідовностями. Запропоноване покращення алгоритму знаходження подібних патернів у часових послідовностях з точки зору продуктивності обчислень.

Створено два Java-додатки для знаходження всіх схожих підпослідовностей у часових рядах, які експериментально підтверджують перевагу у швидкодії запропонованого алгоритму порівняно з класичним та перевагу у швидкодії з існуючою імплементацією SAX – jMotif.

ПЕРЕЛІК ПОСИЛАНЬ

1. Славко О.Г. Інформаційна технологія керування перевантаженнями в мультисервісних телекомунікаційних мережах / Славко О.Г. – 2011 – Вісник Кременчуцького національного університету імені Михайла Остроградського – Вип. 2 (67), част. 1. – С. 29-34.
2. Лаврів О.А. Модель системи управління ресурсами мультисервісних мереж в умовах самоподібності трафіку / Лаврів О. А. – 2012 – Вісник Національного Університету «Львівська політехніка». Радіоелектроніка та телекомунікації. – С. 165-172
3. Черніхівський Є.М., Червенець В. В., Білик О. Б. Визначення часових параметрів обслуговування потокового трафіку з пріоритетними класами / Черніхівський Є.М. – Вісник Національного університету «Львівська політехніка». – 2011. – № 705 : Радіоелектроніка та телекомунікації. – С. 167-170.
4. Rezaul K., Grout V., A Survey of Performance Evaluation and Control for Self-Similar Network Traffic, Proceedings of the Second International Conference on Internet Technologies and Applications (ITA), 4-7 September 2007, pp. 514-524.
5. Samorodnitsky G., Long Range Dependence, Foundations and Trends in Stochastic Systems, 1 (3) (2007) 163-257.
6. Zhang G., Xie G., Yang J., Zhang D., Self-similar characteristic of traffic in current metro area network, Proceedings of the 15th IEEE Workshop on Local and Metropolitan Area Networks, 10-13 June 2007, pp. 176-181.
7. Veitch D., Hohn N., Abry P., Multifractality in TCP/IP traffic: the case against, Computer Networks, 48 (3) (2005) 293-313.
8. Toral-Cruz, Homero, Pathan, Al-Sakib Khan, Pacheco, Julio Cesar Ramirez Accurate modeling of VoIP traffic QoS parameters in current and future networks with multifractal and Markov models – Mathematical and Computer Modelling 57 , no. 11-12 (2013): 2832-2845.
9. Pacheco-Sanchez S., Casale G., Scotney B., McClean S., Parr G., Dawson S. Markovian Workload Characterization for QoS Prediction in the Cloud – 2011 IEEE 4th International Conference on Cloud Computing. 147154 (2011).
10. Islam S., Keung J., Lee K., Liu A.: Empirical prediction models for adaptive resource provisioning in the cloud. Future Generation Computer Systems. 28, 155162 (2012).
11. Mansouri, Taha, Nabavi, Ali, Ravasan, Ahad Zare and Ahangarbahar, Hamid. «A practical model for ensemble estimation of QoS and QoE in VoIP services via fuzzy inference systems and fuzzy evidence theory.» Telecommunication Systems 61 , no. 4 (2016): 861-873.
12. Ролик А.И. Оценка качества предоставления мультимедийных сервисов с использованием нейросетевого классификатора / Ролик А.И., Галушко Д.А., Барна В.В., Томашук А.В., Ясочка М.В. – Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. – К.: Століття +, 2015. – №63. – С. 25–30.
13. Тітов Д.С. Моніторинг соціальних мереж в системі реального часу / Тітов Д.С., Дорошенко А.Ю. – К.: Наукова дискусія: теорія, практика, інновації, 2015. – С. 93-96.
14. Lingfen S. Speech quality prediction for voice over Internet protocol networks: PhD diss.: 2004 / Lingfen Sun. – University of Plymouth, UK, 2004. – 218pp.
15. Lin, J., Keogh, E., Lonardi, S. & Chiu, B. (2003) A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. San Diego, CA. June 13.
16. Lin, J., Keogh, E., Lonardi, S. and Patel P. Finding motifs in time series. In Proc. 2nd Workshop on Temporal Data Mining, 2002. pp. 214-221.
17. Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S. Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases. In proceedings of ACM SIGMOD Conference on Management of Data. Santa Barbara, CA, 2001. pp. 151-162.
18. GitHub (2017) Time series symbolic discretization with SAX [Online] Available from <https://github.com/jMotif/SAX> [Accessed: 12 May 2017]