

Д. М. Росінський, К. О. Свірщевський

Харківський національний технічний університет радіоелектроніки, Харків, Україна

ОПТИМІЗАЦІЯ ЗБОРУ ОПЕРАТИВНИХ ДАНИХ В ПІРИНГОВІЙ МЕРЕЖІ МЕДИЧНОЇ УСТАНОВИ

Анотація. **Актуальність.** В сучасних медичних інформаційних системах обсяг даних, (наприклад, фізіологічних показників, зібраних від пацієнтів), є величезним, оскільки аномальні та нормальні дані збираються разом. Це призводить до затримки надання допомоги невідкладним пацієнтам. Щоб вирішити цю проблему, потрібні оптимальні схеми збору оперативних даних для коротко- і довгострокового прогнозування. При цьому доцільно використовувати пріоритетизацію і фільтрацію даних, які безперервно збираються з переносних датчиків на пацієнтах і надходять до інформаційної системи медичної установи. **Метою даної роботи** є розробка моделі інформаційної системи медичної установи на основі пірингової мережі з використанням ефективних методів збору оперативних даних. **Об'єктом дослідження** є автоматизований процес збору і обробки масивів даних, що відображають стан здоров'я пацієнта медичної установи. **Предметом дослідження** є методи та алгоритми збору і обробки оперативних даних, що характеризують стан пацієнта медичної установи, для формування короткострокового прогнозу. **Результати.** Запропоновані нові алгоритми для оптимізації процесів збору і обробки даних шляхом введення критерія терміновості пацієнтів, що сприятиме зменшенню обсягу даних, які необхідно передати, скороченню часу очікування в постановці діагнозу. Для оцінки терміновості пацієнтів прийняті загальні клінічні критерії. **Висновок.** Запропоновано модель інформаційної системи медичної установи, яка ефективна при зборі даних і оптимізує порядок їхньої обробки в прогнозуванні. Досягнуто зменшення обсягів медичних даних, які надходять від пацієнтів, і налаштування часу очікування для даних, що потрібні в прогнозуванні, на основі різних пріоритетів.

Ключові слова: медична інформаційна система, вектор ознак, класифікація часових рядів, пірингова мережа, пріоритетизація даних.

Вступ

Удосконалення інформаційних систем в сфері охорони здоров'я пов'язане з підвищенням ефективності збору даних і оптимізацією потоків їхньої обробки [1, 2]. Основною задачею стає впровадження пріоритетизації даних на основі терміновості стану пацієнтів, що надасть можливість регулювати час очікування даних перед прогнозуванням. Отже завданнями є оцінка пріоритету на основі послідовностей даних у часі. Разом з тим, виникає інша задача, яка полягає в пришвидшенні передачі оперативних даних, що характеризують поточний стан пацієнтів. Розв'язання цієї задачі можливе за рахунок зменшення обсягу даних, що збираються. Проте, намагаючись зібрати меншу кількість даних, для точного діагнозу необхідно надати клініцистам достатню інформацію. В різних дослідженнях [3-6] для прогнозування аномалій враховуються численні фізіологічні показники (наприклад, артеріальний тиск, частота серцевих скорочень, частота дихання, рівень сатурації, тощо). Але більшість з цих досліджень і досі знаходяться на теоретичному рівні і непридатні для широкого використання. Для визначення того, які дані потрібно передати, окремі автори [7] пропонують до використання приховану марковську модель (НММ), фокусуючись, в основному, на оцінці особистого стану на основі спеціальних правил. Для того, щоб вирішити проблему зберігання та обробки даних, пов'язану з їхньою величезною кількістю, пропонується [8] використовувати хмарну інфраструктуру. Цікавою є пропозиція [3, 7] використовувати мобільний пристрій для безперервного збирання та передачі необроблених фізіологічних показників на сервер. Ще один практичний приклад – апарат BioSign [9], який може мінімізувати імовірність виникнення критичної клінічної ситуації, але не здатен формувати прогноз.

Недоліком наведених вище систем є те, що всі вони обробляють медичні показники пацієнтів за принципом FIFO. Однак у реальних клінічних випадках, щоб визначити оптимальний порядок надання медичних послуг, терміновість пацієнта зазвичай розглядається на початку. Існує приклад [10] ефективної системи сортування, яка використовується у відділенні невідкладної допомоги під час прийняття рішення про скерування пацієнтів до відділення інтенсивної терапії. Отже, під час обробки медичних даних пацієнтів в інформаційній медичній системі потрібно, щоб черговість надходження даних на обробку відображала терміновість стану пацієнтів.

Іншим аспектом дослідження є реалізація концепції машинного навчання з використанням моделі взаємодії типу “Peer-to-Peer” (P2P) [11]. Відомо, що використання централізованих моделей пов'язане із значними витратами на зв'язок через обмін величезними обсягами особистих життєво важливих медичних показників з метою навчання. Використання децентралізованої моделі P2P створює основу для розподіленого обміну знаннями між медичними установами і дозволяє розробити швидку систему машинного навчання, за допомогою якої кожна медична установа може підвищити точність діагностики, вивчаючи знання, отримані від інших установ, і може оптимізувати потік обробки даних пацієнтів на основі різних значень терміновості.

Зважаючи на стан і результати досліджень різних авторів, можна зробити висновок про актуальність питання підвищення ефективності передачі даних і оптимізації потоку даних на основі пріоритетів.

Метою роботи є розробка моделі інформаційної системи медичної установи на основі пірингової мережі з використанням ефективних методів збору оперативних даних.

Основна частина

Модель інформаційної системи медичної установи на основі пірингової мережі продемонстровано на рис. 1. Можна виділити дві частини системи: клієнт для збору даних і сервер для аналізу даних.

Клієнт використовує мобільні пристрої (смартфон, смарт-годинник, тощо) для постійного збирання даних, що відображають фізіологічні показники пацієнтів. Ці дані надходять від різноманітних датчиків на тілі пацієнта. Клієнт також здійснює оцінку пріоритету даних на основі певного часового вікна (наприклад, 5 хвилин). Пріоритет даних відповідає терміновості стану пацієнта: більш «важкий» пацієнт має вищий пріоритет даних. Крім того, ці медичні дані в межах часового вікна з їхніми пріоритетами становлять фрейми показників (ФП). Задачею клієнта є відбір фреймів, які потрібно передати на сервер, виходячи з їхніх пріоритетів.

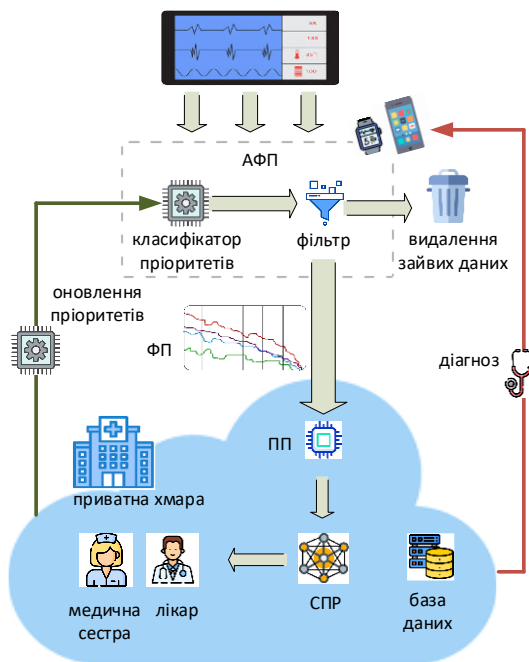


Рис. 1. Модель інформаційної системи медичної установи

Сервер призначений для формування точного короткочасного прогнозу стану здоров'я пацієнтів на основі їхніх фізіологічних показників, отриманих від клієнта. Порядок обробки даних пацієнтів оптимізується за допомогою пріоритетів, що гарантує отримання рішення щодо діагнозу у відповідний час залежно від терміновості пацієнтів. Крім того, завдяки накопиченню даних про пацієнтів, сервер може надати більш точне прогнозування різних медичних станів за допомогою навчання на основі використання пірингової мережі. Основним компонентом клієнта є аналізатор фреймів показників (АФП), який містить два основні компоненти: класифікатор пріоритетів для оцінки даних і фільтр для відбору даних.

Класифікатор пріоритетів використовує метод машинного навчання (МН) для автоматичної класифікації ФП за різними пріоритетами. Пріоритети відіграють важливу роль в оптимізації: класифікатор МН

навчається за допомогою зразків з 4 фізіологічних показників, а мітки навчальних даних ідентифікуються алгоритмом оцінки даних в реальних умовах.

Фільтр допомагає системі ефективніше збирати медичні дані пацієнтів з урахуванням різних пріоритетів. Слід зазначити, що аномальні дані є більш значущими, ніж нормальні, оскільки відповідні діагностичні рішення повинні формуватися під час виявлення аномалій. Таким чином, фільтр націлений на аномальні дані, які мають вищий пріоритет. При цьому потрібна гарантія повноти медичної інформації, яка передається на сервер для діагностики клініцистами.

Сервер містить три компоненти: базу даних для зберігання даних, процесор пріоритетів (ПП) для керування чергою прогнозування та систему прийняття рішень (СПР) для прогнозування. База даних використовується для зберігання записів ФП від різних пацієнтів. Ці записи можна використовувати для встановлення довгострокових діагнозів пацієнтам і видобутку певних корисних медичних знань.

ПП встановлює порядок прогнозування на основі пріоритету даних, що сприяє скороченню часу очікування термінових даних пацієнтів. Довгострокові пріоритети даних визначаються на основі пріоритетів ФП. Наприклад, якщо потрібно обчислити пріоритет даних за годину, потрібні пріоритети 12 часових рядів ФП із використанням 5-хвилинного вікна.

СПР є машинним класифікатором, який може точно прогнозувати стан здоров'я пацієнтів, використовуючи медичні дані, зібрані від пацієнтів у режимі реального часу. З метою забезпечення відповідності змінам у клінічному середовищі (наприклад, появи нових захворювань), для СПР передбачено Р2Р-навчання.

Робота клієнта реалізована у три етапи.

1. Мобільний пристрій постійно збирає дані з датчиків на тілі пацієнта. Вони відображають значення таких фізіологічних показників, як частота дихальних рухів (ЧДР), рівень сатурації киснем (SpO_2), частота серцевих скорочень (ЧСС) та систолічний артеріальний тиск (САТ).

2. АФП у пристрої обробляє дані в межах часових вікон (наприклад, 10 хвилин). Дані у вікні визначаються як ФП. АФП виконує класифікацію пріоритету ФП і фільтрацію ФП.

3. Після фільтрації АФП передає фрейми ФП з індикаторами пріоритетів на сервер медичного закладу.

На сервері обробка даних виконується в 3 етапи.

1. Після отримання достатньої кількості ФП протягом певного вікна спостереження (наприклад, 60 хвилин) сервер групує їх разом і обчислює загальний пріоритет, враховуючи всі ФП в інтервалі спостереження. Далі сервер розміщує ці згруповані дані в списку очікування на основі пріоритету. Вищий пріоритет має вищий індекс, що означає менше часу на обробку.

2. СПР прогнозує медичні стани за допомогою згрупованих ФП і надсилає дані клініцистам після виявлення аномальної клінічної події. Потім пацієнту надаються правильні діагностичні рішення.

3. ФП зберігаються в базі даних для резервного копіювання та подальшого видобутку знань. Дані також використовуються для оновлення класифікатора пріоритетів.

Як показано в табл. 1, бали фізіологічним показникам призначаються на основі їхніх порогових значень. Наприклад, якщо значення ЧДР становить від 9 до 11, то йому надається оцінка 1. Важливо зауважити, що всі значення є цілими. Оцінка 0 означає, що значення знаходиться в межах норми.

Таблиця 1 – Оцінка фізіологічних показників

Оцінка	Значення показника			
	ЧДР, 1/хв.	SpO ₂ , %	ЧСС, 1/хв.	САТ, мм рт./ст.
-3	< 8	< 91	< 40	< 90
-2	-	91-92	-	91-100
-1	9- 11	94-95	41-50	101-110
0	12-20	> 96	51-90	110-120
+1	-	-	91-110	121-170
+2	21-24	-	111-130	171-219
+3	> 25	-	> 131	> 220

Загальний бал усіх показників класифікується за 4 пріоритетами (табл. 2). Критерії класифікації базуються на часі очікування. Зокрема, якщо загальна оцінка дорівнює 0, система розглядає можливість відхилення даних, оскільки всі життєво важливі показники знаходяться в межах норми. Діапазон кожного пріоритету використовується для оцінки послідовної групи пріоритетів, що пояснюється далі.

Таблиця 2 – Визначення пріоритету даних за оцінкою фізіологічних показників

Пріоритет	Мітка	Час очікування	Діапазон	Загальний рахунок
1	норма	-	0	0
2	увага	<1 год.	(0,1]	1-3
3	невідкладна	< 0,5 год.	(1,2]	4-6
4	екстрена	0	(2, 3]	≥ 7 або оцінка +3/-3

Основна роль класифікатора пріоритетів у АФП полягає в позначенні ФП відповідним пріоритетом даних. Припустимо, що дискретна послідовність даних (D) тривалості часу (T) розбивається на L вікон однакового розміру. У кожному вікні є $N = D/L$ зразків. Тоді дані часового ряду можна розглядати як послідовність ФП $D_1(t), D_2(t), \dots, D_L(t)$. Наприклад, якщо в системі використовується 10-хвилинне вікно, 60-хвилинну серію даних можна розділити на послідовність із 6 ФП. Щоб визначити пріоритет ФП, розраховують середнє значення кожного показника із зразків у вікні:

$$f = \sum_{l=1}^N D_l(t) / N. \tag{1}$$

Середні значення усіх фізіологічних показників групуються у вектор ознак $(f_1, f_2, f_3, \dots, f_c)$, де c – загальна кількість показників. Оскільки передбачено обробку лише цілих чисел, усі дробові значення компонент вектора ознак округлюються до найближчого цілого значення. Після виділення ознак ФП класифікатор пріоритетів може визначити його пріоритет, використовуючи будь-які види алгоритмів машинного навчання.

У запропонованій системі фільтр у АФП виділяє значущі дані на основі пріоритету, що сприяє зменшенню обсягу даних. Пріоритет 0 означає, що всі показники належать до нормального діапазону. Якщо в буфері присутній аномальний ФП, усі наступні ФП мають бути взяті до уваги (рис. 2). Коли новий ФП надходить у АФП кожного періоду часу, вікно буфера передачі зсувається ліворуч на крок, що відповідає часовому інтервалу t . Можливі три ідентифікатори, якими позначається ФП: надсилання, очікування та відхилення. Якщо пріоритет ФП у буфері перевищує 0, всі ці 6 ФП надсилаються на сервер. Якщо пріоритети нового ФП та решти 5 наступних дорівнюють 0, він позначається як очікуваний. Коли буфер зміщується від очікуючого ФП, той відкидається. Таким чином, алгоритм коригує процес збору даних, фокусуючись на аномальних ФП.

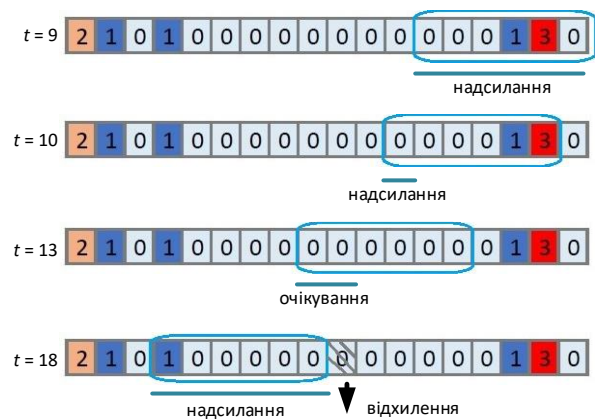


Рис. 2. Приклад передачі даних

Для визначення загального пріоритету ФП використовуються всього L пріоритетів фрагментів послідовностей даних. Замість усереднення всіх пріоритетів, використовуються вагові коефіцієнти часу:

$$w_i = T_i^2 / \sum_{i=1}^L T_i^2; T_i = i \times W. \tag{2}$$

Позначимо множину всіх пріоритетів як P . Тоді $P = (p_1, p_2, \dots, p_L)$. Загальний пріоритет p' безперервних ФП обчислюється наступним чином:

$$p' = P \times W = \sum_{i=1}^L p_i \times w_i. \tag{3}$$

Система поміщає всі видобуті функції послідовних ФП у список очікування на основі їхнього пріоритету. Для обробки вибираються дані з найвищим пріоритетом. У кожному ФП виділяється 5 ознак, якими є середнє значення, стандартне відхилення, медіана, кількість тенденцій зростання та тенденцій до зменшення.

Критеріями ефективності запропонованих рішень можуть бути: точність класифікації пріоритетів; продуктивність збору даних; середній час очікування.

Таблиця 3 – Матриця помилок класифікації

Пріоритет	1	2	3	4
1	445	136	10	1
2	4	248	18	8
3	1	66	380	63
4	0	0	42	378

Використано реальні медичні дані з МІМІС-II [11]. Щоб збалансувати набір даних, випадковим чином вибираються 1500 зразків кожного пріоритету. Набір даних нормалізується за допомогою лінійного перетворення z-показника. Була отримана оцінка точності класифікації пріоритетів. Для навчання використано 70% вибірок, решта – для тестування (табл. 3). Крім загальної матриці помилок, можуть застосовуватися інші показники точності для кожного пріоритету.

Висновки

Досліджено автоматизований процес збору і обробки масивів даних, що відображає стан здоров'я пацієнта медичної установи. Запропонована модель ефективна при зборі даних, оптимізує порядок обробки даних в прогнозуванні. Вона сприяє зменшенню обсягів медичних даних, зібраних від пацієнтів, фільтруючи звичайні дані, і може налаштувати час очікування даних перед прогнозуванням на основі різних пріоритетів. Використання децентралізованої моделі P2P надає можли-

вість розподіленого обміну знаннями між різними медичними установами і дозволяє розробити швидку систему машинного навчання, за допомогою якої кожна медична установа може підвищити точність діагностики, вивчаючи знання, отримані від інших установ, і може оптимізувати потік обробки даних пацієнтів.

Отже, отримано модель системи з оцінкою пріоритету даних, що оптимізує збір даних і прогнозування щодо розміру даних і часу очікування. Розроблено новий алгоритм оцінки пріоритету даних, який може представляти терміновість пацієнтів. Завдяки пріоритетизації даних, у запропонованій системі не тільки збирається менше даних від пацієнтів, але також надається повна медична інформація, необхідна для довгострокового встановлення точного діагнозу.

Розглядаючи різні розподіли пріоритету даних у реальному випадку, експериментальні результати свідчать про те, що запропоновані рішення підвищують ефективність збору даних і оптимізують час очікування відповідно до терміновості пацієнтів.

СПИСОК ЛІТЕРАТУРИ

1. Frick N., Mirbabaie M., Stieglitz S., and Salomon J. (2021). "Maneuvering through the stormy seas of digital transformation: the impact of empowering leadership on the AI readiness of enterprises." *Journal Of Decision Systems*, 1-24.
2. Bohr A., and Memarzadeh K. (2020). "Artificial intelligence in healthcare." San Diego: Elsevier Science & Technology.
3. Forkan A. R. M., and Khalil I. (2017). "Peace-home: Probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring." *Pervasive and Mobile Computing*, 38:296-311.
4. Li H., and Boulanger P. (2020). "A Survey of Heart Anomaly Detection Using Ambulatory Electrocardiogram (ECG)." *Sensors*, 20(5), 1461. doi: 10.3390/s20051461
5. Cao H., Eshelman L., Chbat N., Nielsen L., Gross B., and Saeed M.. (2008). "Predicting icu hemodynamic instability using continuous multiparameter trends." In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 3803-3806.
6. Salem O., Liu Y., Mehaoua A., and Boutaba R.. (2014). "Online anomaly detection in wireless body area networks for reliable healthcare monitoring." *IEEE J. Biomedical and Health Informatics*, 18(5):1541-1551.
7. Jiang P., Winkley J., Zhao C., Munnoch R., Min G., and Yang L. T. (2016). "An intelligent information forwarder for healthcare big data systems with distributed wearable sensors." *IEEE Systems Journal*, 10(3):1147-1159.
8. Holm S., Stanton C., and Bartlett B. (2021). "A New Argument for No-Fault Compensation in Health Care: The Introduction of Artificial Intelligence Systems." *Health Care Analysis*. doi: 10.1007/s10728-021-00430-4
9. Tarassenko L., Hann A., Patterson A., Braithwaite E., Davidson K., Barber V., and Young D. (2005). "Biosign: multi-parameter monitoring for early warning of patient deterioration." pages 71-76.
10. Xie R., Khalil I., Badsha S., and Atiquzzaman M. (2018). "Fast and peer-to-peer vital signal learning system for cloud-based healthcare." *Future Generation Computer Systems*, 88:220-233.
11. Saeed M., Villarroel M., Reisner A. T., Clifford G., Lehman L.-W., Moody G., Heldt T., Kyaw T. H., Moody B., and Mark R. G.. (2019). "Multiparameter intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database." *Critical care medicine*, 39(5):952.

Received (Надійшла) 23.03.2023

Accepted for publication (Прийнята до друку) 17.05.2023

Optimizing operational data collection in a medical institution's peer-to-peer network

Dmytro Rosinskiy, Kyryl Svirshchevskyi

Abstract. Topicality. In modern medical information systems, the volume of data (for example, physiological parameters collected from patients) is enormous, as abnormal and normal data are collected together. This leads to delays in providing care to emergency patients. To solve this problem, optimal operational data collection schemes for short- and long-term forecasting are needed. At the same time, it is advisable to use prioritization and filtering of data that are continuously collected from portable sensors on patients and sent to the information system of the medical institution. **The goal of this work** is to develop a model of the information system of a medical institution based on a peering network using effective methods of collecting operational data. **The object of research** is the automated process of collecting and processing data arrays, which will reflect the state of health of a patient of a medical institution. **The subject of research** is methods and algorithms for collecting and processing operational data characterizing the condition of a patient in a medical institution, for the formation of a short-term prognosis. **Results.** New algorithms are proposed to optimize the processes of data collection and processing by introducing a criterion of urgency for patients, which will contribute to reducing the amount of data that must be transferred, reducing the waiting time for making a diagnosis. Common clinical criteria are used to assess the urgency of patients. **Conclusions.** A model of the information system of a medical institution is proposed, which is effective in collecting data and optimizes the order of their processing in forecasting. Achieved a reduction in the amount of medical data received from patients and the setting of the waiting time for the data required for forecasting.

Keywords: medical information system, feature vector, time series classification, peering network, data prioritization.