

С.В. Поперешняк

БАГАТОКОМПОНЕНТНА МОДЕЛЬ СЕКВЕНЦІЙНОГО АГРЕГОВАНОГО ОЦІНЮВАННЯ КОРОТКИХ ПОСЛІДОВНОСТЕЙ ДЛЯ ВИЯВЛЕННЯ ЗМІН СТАНУ СИСТЕМИ

У статті запропоновано багатокомпонентну модель секвенційного агрегованого оцінювання коротких послідовностей, призначену для виявлення змін стану систем за обмеженого обсягу спостережень. Актуальність дослідження зумовлена необхідністю оперативного аналізу поточкових даних у випадках, коли використання довгих часових рядів, накопичення значних обсягів статистичних даних або навчання складних моделей є недоцільним чи неможливим. Запропонована модель ґрунтується на точних спільних розподілах структурних статистик коротких бінарних послідовностей і передбачає послідовне виконання двох взаємопов'язаних рівнів агрегування. На першому рівні формується локальна структурно-ймовірнісна оцінка окремого короткого фрагмента з урахуванням сукупності взаємопов'язаних характеристик його внутрішньої структури. На другому рівні здійснюється секвенційне агрегування впорядкованої траєкторії локальних оцінок зі збереженням інформації про характер їхньої зміни. Результатом є багатокомпонентна оцінка, що включає характеристики рівня, напрямку зміни, накопичення, тривалості та мінливості структурної нетиповості. Такий підхід дає змогу характеризувати не лише наявність локального відхилення, а й особливості його формування та розвитку в послідовності спостережень. Проведено експериментальне дослідження моделі, яке підтвердило можливість розрізнення коротких послідовностей з однаковою кількістю окремих подій, але різною внутрішньою структурою, а також відокремлення одиничних локальних відхилень від стійких і спрямованих структурних змін. Практичну апробацію виконано на даних IoT-моніторингу стану прісноводної екосистеми. Отримані результати показали можливість ідентифікації різних фаз зміни стану контрольованого процесу та підтвердили практичну придатність моделі для потокового моніторингу, раннього виявлення структурних змін і формування інтерпретованих ознак для подальших задач класифікації та прогнозування.

Ключові слова: короткі послідовності, бінарні послідовності, структурно-ймовірнісне оцінювання, секвенційне агрегування, багатокомпонентна модель, виявлення змін стану, структурна нетиповість, IoT-моніторинг, поточкові дані, прогнозування.

S. Popereshnyak

MULTICOMPONENT MODEL FOR SEQUENTIAL AGGREGATED ASSESSMENT OF SHORT SEQUENCES FOR SYSTEM STATE CHANGE DETECTION

This paper proposes a multicomponent model for sequential aggregated assessment of short sequences designed to detect system state changes under a limited number of observations. The relevance of the study is driven by the need for rapid analysis of streaming data in cases where the use of long time series, the accumulation of large volumes of statistical data, or the training of complex models is impractical or infeasible. The proposed model is based on exact joint distributions of structural statistics of short binary sequences and comprises two interrelated levels of aggregation. At the first level, a local structural-probabilistic assessment of an individual short fragment is formed by considering a set of interrelated characteristics of its internal structure. At the second level, sequential aggregation of an ordered trajectory of local assessments is performed while preserving information about the nature of their changes. The resulting multicomponent assessment incorporates characteristics of the level, direction of change, accumulation, duration, and variability of structural atypicality. This approach makes it possible to characterize not only the presence of a local deviation but also the specific features of its formation and evolution throughout the sequence of observations. An experimental study of the model was conducted and confirmed its ability to distinguish between short sequences containing the same number of individual events but exhibiting different internal structures, as well as to differentiate isolated local deviations from persistent and directional structural changes. The model was practically validated using IoT monitoring data from a freshwater ecosystem. The obtained results demonstrated the ability to identify different phases of changes in the state of the monitored process and confirmed the practical applicability of the proposed model to streaming monitoring, early detection of structural changes, and the generation of interpretable features for subsequent classification and forecasting tasks.

Keywords: short sequences, binary sequences, structural-probabilistic assessment, sequential aggregation, multicomponent model, system state change detection, structural atypicality, IoT monitoring, streaming data, forecasting.

Вступ

Розвиток сучасних інформаційних технологій моніторингу та керування супроводжується зростанням кількості інформаційних, кіберфізичних, сенсорних та IoT-систем, стан яких необхідно оцінювати безперервно або через малі інтервали часу. Своєчасне виявлення зміни стану таких систем є важливою умовою забезпечення надійності їх функціонування, попередження критичних режимів та підтримки ухвалення рішень. Водночас у багатьох практичних задачах оцінювання необхідно здійснювати за обмеженою кількістю останніх спостережень, не очікуючи накопичення вибірки значного обсягу. Це зумовлює актуальність методів аналізу коротких послідовностей, орієнтованих на оперативне виявлення початкових проявів зміни стану системи.

Особливість таких задач полягає в тому, що зміна стану не завжди супроводжується різким виходом контрольованого параметра за допустимі межі. На початкових етапах вона може проявлятися у зміні частоти виникнення відхилень, характеру їх чергування, тривалості окремих станів або взаємного розташування подій. Подання результатів спостережень у вигляді коротких бінарних послідовностей дає змогу перейти від аналізу абсолютних значень параметра до дослідження структури подій. Водночас однакова кількість символів певного типу може відповідати суттєво різним варіантам їх розташування, тому використання лише окремої кількісної статистики не забезпечує достатньої інформативності.

Це визначає необхідність одночасного врахування декількох взаємопов'язаних структурних характеристик короткої послідовності та статистичної допустимості їх спільної реалізації. За малого обсягу спостережень такий підхід є особливо важливим, оскільки зміни окремих статистик можуть бути наслідком як зміни стану системи, так і природної варіативності випадкової послідовності. Тому структурний

опис доцільно поєднувати з імовірнісним оцінюванням відповідних комбінацій характеристик.

Водночас багатокомпонентна оцінка окремого короткого фрагмента характеризує систему лише в межах локального інтервалу. Для оперативного моніторингу необхідно встановити, чи є виявлене відхилення випадковою флуктуацією або складовою стійкої зміни. Це потребує аналізу впорядкованої послідовності локальних оцінок та врахування не лише їхнього рівня, а й напрямку зміни, накопичення, тривалості та мінливості. Поєднання структурно-імовірнісного опису коротких фрагментів із секвенційним агрегуванням їхніх оцінок дозволяє одночасно враховувати внутрішню структуру подій і динаміку її зміни.

Практично значущою сферою застосування такого підходу є IoT-моніторинг, де сенсорні вузли формують потоки значень параметрів контрольованих об'єктів і середовища. Зміна режиму функціонування може проявлятися у поступовому збільшенні частоти відхилень або зміні характеру їх групування ще до стійкого перевищення критичних рівнів. Отже, **актуальним** є розроблення моделі, яка поєднує багатокомпонентне структурно-імовірнісне оцінювання коротких послідовностей із секвенційним агрегуванням локальних оцінок та забезпечує виявлення стійких змін стану систем за обмеженого обсягу спостережень.

Аналіз останніх досліджень і публікацій

Виявлення аномалій і змін стану систем за часовими послідовностями є актуальним напрямом сучасного аналізу даних. Для розв'язання таких задач застосовують статистичні методи, алгоритми машинного та глибокого навчання, зокрема рекурентні нейронні мережі, автоенкодера та трансформерні архітектури [1, 2]. Сучасні дослідження орієнтовані не лише на аналіз окре-

мих значень, а й на врахування часових залежностей, контексту та взаємозв'язків між ознаками.

Значного розвитку набули трансформерні моделі аналізу багатовимірних часових рядів. У Anomaly Transformer [3] для виявлення аномалій використано відмінності асоціативних зв'язків між елементами послідовності, а TranAD [4] поєднує аналіз часових залежностей і багатовимірною представлення даних. Розвиток ідеї комплексного опису часових даних представлений у TFAD [5], де поєднуються часові та частотні характеристики, Anomaly-PTG [6], що враховує часові та міжознакові залежності, та TimesNet [7], у якому формуються багатокомпонентні двовимірні представлення часових закономірностей. Використання багатомасштабного представлення та контрастивного навчання для підвищення чутливості до структурних відмінностей реалізовано у DCdetector [8].

Проведені в останні роки узагальнення [9, 10] свідчать про стійку тенденцію до використання комплексних представлень часових послідовностей та врахування їхньої локальної й глобальної структури. Водночас значна частина сучасних підходів орієнтована на досить великі масиви даних та навчання параметричних моделей. Це обмежує можливості їх використання в задачах оперативного моніторингу, коли зміни стану необхідно виявляти за короткими послідовностями спостережень.

Окремий напрям досліджень пов'язаний зі статистичним аналізом коротких бінарних послідовностей та локальних особливостей розташування їхніх елементів. У роботі [11] запропоновано підхід до перевірки випадковості розташування бітів у локальних сегментах бінарної послідовності на основі статистик входжень визначених ланцюжків. Подальший розвиток цього підходу представлено в [12], де отримано спільні розподіли статистик випадкових бінарних послідовностей, що створює математичні передумови для одночасного ймовірнісного оцінювання декількох взаємопов'язаних структурних характеристик коротких фрагментів.

Для таких послідовностей використання окремої статистичної характеристики

може бути недостатнім, оскільки однаковим її значенням можуть відповідати різні структури розташування подій. Одночасне використання декількох характеристик також потребує врахування їх взаємозалежності та ймовірності спільної реалізації. Крім того, для виявлення стійкої зміни стану важливим є аналіз не лише окремої локальної оцінки, а й динаміки таких оцінок у послідовних інтервалах спостереження. Отже, недостатньо дослідженою залишається задача формування інтерпретованої багатокомпонентної структурно-ймовірнісної оцінки коротких послідовностей та секвенційного агрегування локальних оцінок для виявлення змін стану системи.

Метою дослідження є розроблення та експериментальне дослідження багатокомпонентної моделі секвенційного агрегованого оцінювання коротких послідовностей для виявлення змін стану системи. Для досягнення мети необхідно формалізувати багатокомпонентне структурно-ймовірнісне представлення короткої послідовності; визначити структуру та взаємозв'язки компонентів моделі; розробити процедуру формування й секвенційного агрегування локальних оцінок; визначити принципи інтерпретації отриманих результатів; провести експериментальне дослідження моделі та здійснити її практичну апробацію для виявлення змін стану IoT-системи за сенсорними даними.

Виклад основного матеріалу

Постановка задачі багатокомпонентного оцінювання коротких послідовностей. Розглянемо інформаційну систему моніторингу, в якій стан контрольованого об'єкта характеризується послідовністю спостережень

$$X = \{x_1, x_2, \dots, x_N\}.$$

Залежно від предметної області значення x_i можуть відповідати результатам сенсорних вимірювань, параметрам функціонування технічної або інформаційної системи, зареєстрованим подіям чи іншим спостережуваним характеристикам. Для уніфікації подальшого аналізу результати спостережень перетворюються на бінарну

форму відповідно до заданого правила $g(x)$:

$$b_i = g(x_i) = \begin{cases} 0, & x_i \in \Omega_0, \\ 1, & x_i \in \Omega_1 \end{cases} i = 1, \dots, N,$$

де Ω_0 та Ω_1 — області значень, що відповідають визначеним станам контрольованого параметра. Для сенсорних даних IoT-систем таким способом, наприклад, можуть бути представлені нормальний стан параметра та його відхилення від заданого контрольованого рівня.

У результаті формується бінарна послідовність

$$B = (b_1, b_2, \dots, b_N), b_i \in \{0, 1\}.$$

Оскільки задача полягає в оперативному виявленні зміни стану системи за обмеженим обсягом спостережень, послідовність B аналізується в межах коротких ковзних вікон довжини n . Для моменту t відповідний фрагмент визначається як

$$W_t = (b_{t-n+1}, b_{t-n+2}, \dots, b_t), t = n, \dots, N.$$

Кожний фрагмент W_t характеризується сукупністю структурних статистик. У межах дослідження використовується трикомпонентне структурне представлення

$$K_t = (k_{1,t}, k_{2,t}, k_{3,t}),$$

де $k_{1,t}$, $k_{2,t}$ та $k_{3,t}$ характеризують різні взаємопов'язані властивості внутрішньої структури короткої бінарної послідовності.

Необхідність використання декількох структурних характеристик зумовлена тим, що значення окремої статистики не визначає структуру короткої послідовності однозначно. Два фрагменти можуть мати однакову кількість подій певного типу, але суттєво відрізнятися характером їх розташування. Наприклад, послідовності

$$W_a = (0001000100010001), \\ W_b = (0000000011110000)$$

містять однакову кількість одиничних символів, проте в першій вони розподілені в межах усього інтервалу спостереження, а в другій утворюють локалізовану групу. Отже, якщо k_1 характеризує лише кількість подій певного типу, то

$$k_1(W_a) = k_1(W_b),$$

тоді як за іншими структурними характеристиками

$$(k_2(W_a), k_3(W_a)) \neq (k_2(W_b), k_3(W_b)).$$

Для задач моніторингу така відмінність є суттєвою, оскільки однакова кількість відхилень контрольованого параметра може відповідати різним режимам функціонування системи. Поодинокі розподілені відхилення можуть бути наслідком природної варіативності спостережень, тоді як їхнє локальне групування може характеризувати формування стійкої зміни стану.

Водночас незалежне використання декількох статистик також не забезпечує повного оцінювання структури фрагмента. Значення $k_{1,t}$, $k_{2,t}$ та $k_{3,t}$ отримуються для однієї послідовності W_t , тому необхідно враховувати не лише кожне з них окремо, а й можливість їх спільної реалізації. Відповідно, структурному вектору K_t ставиться у відповідність спільна ймовірність

$$P_t = P\{K_1 = k_{1,t}, K_2 = k_{2,t}, K_3 = k_{3,t}\}.$$

Таке представлення дозволяє оцінювати статистичну допустимість конкретної комбінації структурних характеристик короткої послідовності та створює основу для формування її багатокомпонентної структурно-ймовірнісної оцінки.

Однак для виявлення зміни стану системи оцінювання окремого короткого фрагмента є недостатнім. У процесі надходження нових даних ковзне вікно послідовно зміщується, внаслідок чого формується множина фрагментів

$$W_n, W_{n+1}, \dots, W_N$$

та відповідна послідовність локальних структурно-ймовірнісних оцінок

$$Q_n, Q_{n+1}, \dots, Q_N.$$

Одинична нетипова локальна оцінка може бути результатом випадкової варіативності короткої послідовності. Натомість узгоджена зміна декількох послідовних оцінок може характеризувати стійку трансформацію структури спостережень і, відповідно, зміну стану контрольованої системи. Тому локальні оцінки необхідно розглядати не ізольовано, а з урахуванням послідовності їх формування.

Таким чином, задача багатокompонентного секвенційного оцінювання полягає у реалізації послідовного перетворення

$$W_t \rightarrow K_t \rightarrow Q_t \rightarrow A_t,$$

де W_t – поточна коротка бінарна послідовність; $K_t = (k_{1,t}, k_{2,t}, k_{3,t})$ – її структурне представлення; Q_t – багатокompонентна локальна структурно-ймовірнісна оцінка; A_t – секвенційна агрегована оцінка, що формується з урахуванням результатів аналізу послідовних коротких фрагментів.

Отже, для розв'язання поставленої задачі необхідно визначити склад багатокompонентної локальної оцінки Q_t , встановити спосіб її формування на основі структурних характеристик та їх спільних імовірнісних розподілів, а також визначити механізм секвенційного агрегування локальних оцінок для виявлення зміни стану системи. Відповідні положення становлять основу багатокompонентної моделі секвенційного агрегованого оцінювання коротких послідовностей.

Багатокompонентна модель секвенційного агрегованого оцінювання коротких послідовностей. Для розв'язання поставленої задачі запропоновано багатокompонентну модель секвенційного агрегованого оцінювання коротких послідовностей. Її основою є послідовне перетворення локальних структурних характеристик коротких фрагментів на структурно-ймовірнісні оцінки, а отриманої траєкторії локальних оцінок – на систему характеристик, що відображають динаміку зміни стану досліджуваного процесу.

Модель має два взаємопов'язані рівні агрегування. Перший рівень забезпечує формування інтегральної локальної оцінки окремого короткого фрагмента на основі декількох структурно-ймовірнісних характеристик. Другий рівень забезпечує секвенційне агрегування впорядкованої послідовності локальних оцінок із збереженням інформації про характер їх зміни. Такий поділ відповідає сформованій у дисертаційному дослідженні багаторівневій моделі переходу від локальних статистик до секвенційно агрегованої оцінки.

Для r -го короткого фрагмента W_r формується вектор локальних структурно-ймовірнісних оцінок

$$Q_r = (q_{1,r}, q_{2,r}, \dots, q_{m,r}), q_{j,r} \in [0,1],$$

де $q_{j,r}$ — нормована оцінка структурної не-типності фрагмента за j -ю системою структурних статистик. Вихідною основою для її визначення є точні спільні розподіли статистик коротких послідовностей, що дозволяє враховувати не ізольовані значення структурних характеристик, а статистичну допустимість їх спільної реалізації.

Структурне агрегування локальних оцінок. Для переходу від багатовимірного представлення Q_r до інтегральної локальної характеристики введемо оператор структурного агрегування G :

$$e_r = G(Q_r), e_r \in [0,1].$$

Вибір конкретного оператора G залежить від змісту прикладної задачі та характеру взаємодії структурних характеристик.

За однакової значущості компонентів інтегральна локальна оцінка визначається середнім значенням:

$$e_r = \frac{1}{m} \sum_{j=1}^m q_{j,r}.$$

Якщо окремі структурні характеристики мають різну значущість, використовується зважене агрегування:

$$e_r = \sum_{j=1}^m w_j q_{j,r},$$

за умов $w_i \geq 0, \sum_{j=1}^m w_i = 1$.

Зважене агрегування дозволяє врахувати внесок усіх компонентів, проте допускає їх взаємну компенсацію. Якщо ж значне відхилення хоча б однієї структурної характеристики має бути збережене незалежно від значень інших компонентів, доцільним є використання максимального значення:

$$e_r = \max_{1 \leq j \leq m} q_{j,r}.$$

Таким чином, результатом першого рівня моделі є нормована локальна оцінка e_r , яка узагальнює структурно-ймовірнісні характеристики окремого короткого фрагмента.

Послідовне зміщення ковзного вікна формує впорядковану траєкторію локальних оцінок

$$\mathbf{E} = (e_1, e_2, \dots, e_R),$$

де R — кількість сформованих коротких фрагментів.

Принциповим є те, що у подальшому аналізі зберігається **порядок** локальних оцінок. Наприклад, траєкторії $(0,1;0,2;0,7;0,9)$ та $(0,9;0,7;0,2;0,1)$ мають однакові середні значення, але характеризують протилежні процеси - зростання та зменшення структурної нетиповості. Саме врахування впорядкованої траєкторії локальних оцінок відрізняє секвенційне агрегування від звичайного інтегрального узагальнення.

Для поточного положення r розглядається інтервал секвенційного аналізу довжини s :

$$\mathbf{E}_r^{(s)} = (e_{r-s+1}, e_{r-s+2}, \dots, e_r).$$

На цьому інтервалі визначається багатокомпонентна секвенційно агрегована оцінка, яка включає п'ять взаємодоповнювальних характеристик: **рівень, напрям зміни, накопичення, тривалість і мінливість**. Саме ці компоненти в дисертаційному дослідженні утворюють багатокомпонентний опис структурної динаміки коротких послідовностей.

Компонента рівня. Компонента рівня характеризує загальну структурну нетиповість на поточному інтервалі секвенційного аналізу:

$$L_r = \frac{1}{s} \sum_{i=r-s+1}^r e_i. \quad (1)$$

Зростання L_r відповідає загальному підвищенню структурної нетиповості послідовності. Водночас вплив одиничного короткочасного відхилення зменшується за рахунок агрегування локальних оцінок на інтервалі s .

Компонента напрямку зміни. Для визначення спрямованості структурних змін використовується трендова компонента. Її можна визначити за нормованим коефіцієнтом нахилу лінійної апроксимації локальних оцінок:

$$T_r = \frac{\beta_r}{\beta_{\max}}, -1 \leq T_r \leq 1, \quad (2)$$

де β_r — коефіцієнт нахилу лінійної регресії для значень $\mathbf{E}_r^{(s)}$; $\beta_{\max}$ - нормувальне значення.

При $T_r > 0$ структурна нетиповість має тенденцію до зростання, при $T_r < 0$ — до зменшення, а значення T_r , близьке до нуля, відповідає відсутності вираженої спрямованої зміни.

Компонента накопичення. Для оцінювання накопичення структурних відхилень вводиться контрольний рівень θ . Компонента накопичення визначається як

$$C_r = \frac{1}{s} \sum_{i=r-s+1}^r \max(0, e_i - \theta). \quad (3)$$

На відміну від компоненти рівня, C_r враховує саме частину локальних оцінок, що перевищує встановлений контрольний рівень. Зростання C_r характеризує поступове накопичення структурно нетипових фрагментів.

Компонента тривалості. Нехай d_r — довжина поточної неперервної серії локальних оцінок, для яких $e_i > \theta$. Тоді нормована компонента тривалості визначається як

$$D_r = \frac{d_r}{s}, 0 \leq D_r \leq 1. \quad (4)$$

Компонента D_r дозволяє розрізняти декілька ізольованих перевищень та неперервну серію структурно нетипових фрагментів. Тому вона характеризує стійкість виявленої зміни.

Компонента мінливості. Мінливість локальних оцінок визначається їх відхиленням від поточного середнього рівня:

$$V_r = \frac{1}{V_{\max}} \sqrt{\frac{1}{s} \sum_{i=r-s+1}^r (e_i - L_r)^2}, \quad (5)$$

де $V_{\max}$ - нормувальне значення.

Компонента V_r характеризує стабільність траєкторії локальних оцінок і дозволяє відрізняти стійкий режим структурної зміни від коливального.

Таким чином, результатом другого рівня моделі є векторна секвенційно агрегована оцінка

$$\mathbf{A}_r = (L_r, T_r, C_r, D_r, V_r).$$

Функціональне призначення компонентів моделі узагальнено в табл. 1.

Таблиця 1.

Компоненти багатокомпонентної
секвенційно агрегованої оцінки

Компо- нента	Позна- чення	Характеристика
Рівень	L_r	загальний рівень стру- ктурної нетиповості
Напря- м зміни	T_r	спрямованість розви- тку структурних змін
Накопи- чення	C_r	інтенсивність накопи- чення відхилень
Трива- лість	D_r	стійкість і неперер- вність відхилення
Мінли- вість	V_r	стабільність або коли- вальність процесу

Компоненти A_r мають взаємодоповнювальний характер. Високе значення L_r не визначає напрям зміни; додатне значення T_r не свідчить про тривалість відхилення; значне C_r може бути сформоване декількома розділеними перевищеннями, тоді як D_r характеризує їх неперервність; V_r дозволяє додатково визначити стабільний або коливальний характер процесу. Саме поєднання цих характеристик забезпечує багатокомпонентний опис структурної динаміки.

Загальне перетворення в межах запропонованої моделі можна представити як

$$Q_r \rightarrow e_r \rightarrow E_r^{(s)} \rightarrow A_r,$$

де перехід $Q_r \rightarrow e_r$ відповідає **структурному агрегуванню** локальних оцінок одного короткого фрагмента, а перехід $e_r \rightarrow E_r^{(s)} \rightarrow A_r$ – **секвенційному агрегуванню** впорядкованої траєкторії локальних оцінок.

Структурно запропонована модель реалізує два взаємопов'язані рівні агрегування. На першому рівні для кожного короткого фрагмента послідовності визначається сукупність структурних статистик, на основі яких формуються локальні структурно-ймовірнісні оцінки та виконується їхнє структурне агрегування з отриманням інтегральної локальної оцінки e_r . На другому рівні послідовність таких оцінок розглядається як упорядкована траєкторія, для якої визначаються характеристики рівня, напрямку зміни, накопичення, тривалості та мінливості. Узагальнену структурну схему багатокомпонентної моделі секвенційного агрегованого оцінювання наведено на рис. 1.

Як видно з рис. 1, перший рівень моделі реалізує перехід від короткого фрагмента W_r до локальної інтегральної оцінки e_r через визначення структурних статистик K_r , формування вектора структурно-ймовірнісних оцінок Q_r та їх агрегування оператором G . Другий рівень використовує не окреме значення e_r , а впорядковану сукупність локальних оцінок $E_r^{(s)}$, що дозволяє врахувати характер їхньої зміни в межах заданого секвенційного інтервалу.

На цьому рівні паралельно формуються п'ять взаємодоповнювальних компонентів: рівень L_r , напрям зміни T_r , накопичення C_r , тривалість D_r та мінливість V_r . Сукупність цих компонентів утворює багатокомпонентну секвенційно агреговану оцінку.

Таким чином, модель забезпечує перехід від локального структурно-ймовірнісного опису окремого короткого фрагмента до структурно-динамічного представлення послідовності спостережень. У той же час зберігається інформація не лише про рівень структурної нетиповості, а й про напрям, інтенсивність накопичення, стійкість та мінливість її розвитку, що створює основу для розрізнення сценаріїв зміни стану системи.

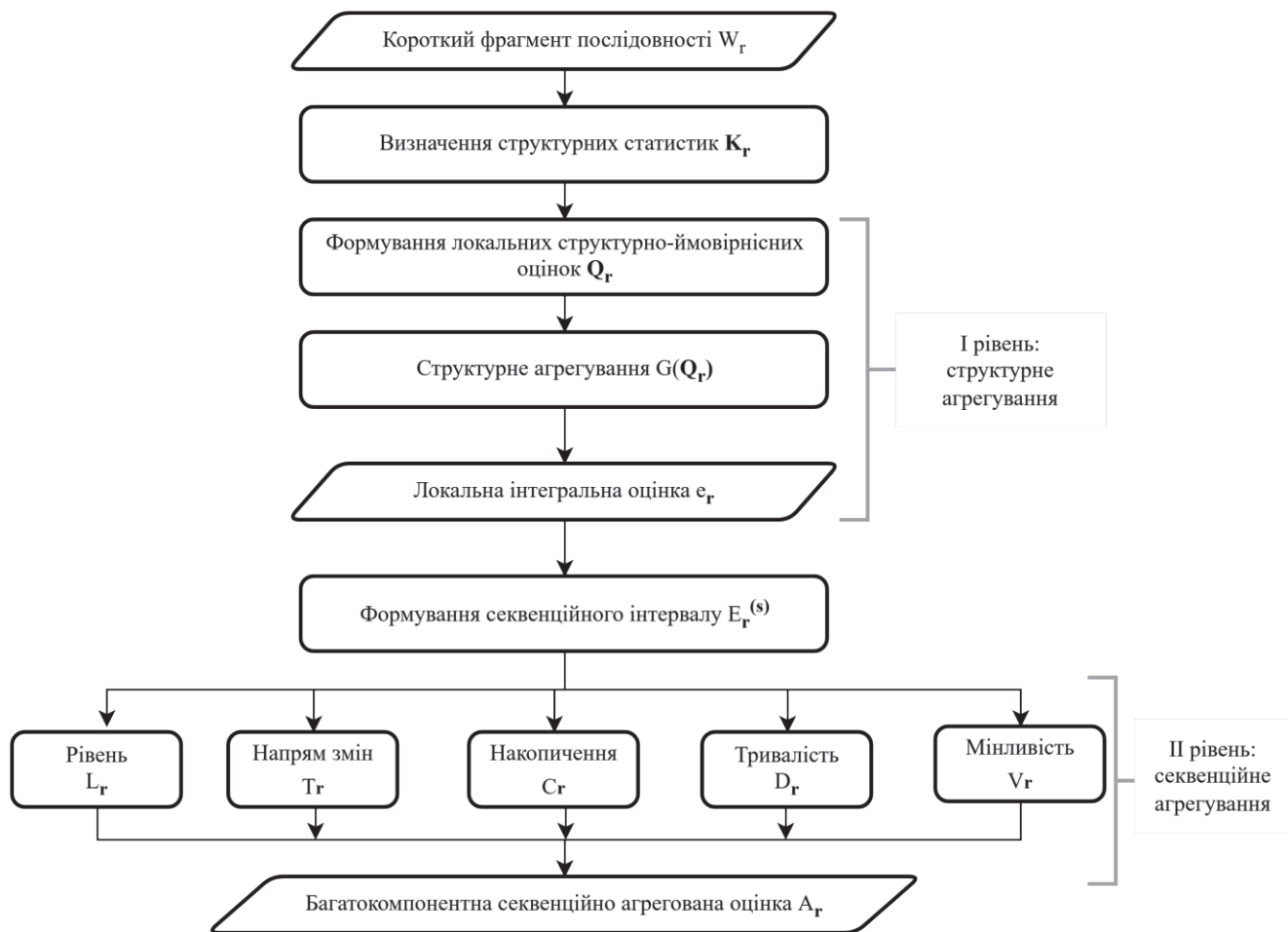


Рис. 1. Структурна схема багатокомпонентної моделі секвенційного агрегованого оцінювання коротких послідовностей

Експериментальне дослідження моделі

Експериментальне дослідження проведено у два етапи. На першому перевірялась здатність багатокомпонентного структурно-ймовірнісного оцінювання розрізняти короткі послідовності з однаковою кількістю подій, але різним характером їх розташування. На другому досліджувалась здатність секвенційного агрегування розрізняти стабільний режим, одиничне структурне відхилення та стійку зміну структури.

Для забезпечення відтворюваності експерименту використано бінарні послідовності довжини $n=16$ за умови рівноймовірної появи символів $p=q=0,5$. Такий випадок відповідає використаному в математичному апараті підходу до дослідження малих вибірок і точних спільних розподілів структурних статистик.

Дослідження багатокомпонентного оцінювання. Для першого етапу сформовано три послідовності довжини $n = 16$, кожна з яких містить чотири одиничні символи:

$$W_1 = (1000100010001000),$$

$$W_2 = (0011000011000000),$$

$$W_3 = (0000001111000000).$$

Таким чином, за кількістю одиничних подій усі три послідовності однакові. Водночас характер їх розташування принципово відрізняється: у W_1 одиниці ізольовані, у W_2 формуються дві локальні групи, а у W_3 – одна неперервна група з чотирьох одиниць.

Як структурні характеристики використано кількість входжень ланцюжків 11, 111 та 101: $k_1 = \eta(11)$, $k_2 = \eta(111)$, $k_3 = \eta(101)$.

Використання числа входжень двота триланцюжків відповідає прийнятому математичному апарату точних спільних розподілів. Зокрема, для послідовностей скінченної довжини такі статистики дозволяють кількісно характеризувати особливості локального розташування бінарних символів.

Для кожного структурного вектора (k_1, k_2, k_3) шляхом повного перебору 65536 можливих послідовностей визначено точну ймовірність P його реалізації. Для переходу до єдиної шкали структурної нетиповості використано нормовану оцінку q , значення якої зростає зі зменшенням типовості відповідної структурної конфігурації.

Результати наведено в табл. 2.

Таблиця 2.

Результати структурно-ймовірнісного оцінювання послідовностей різної структури

Послідовність	$k_1 = \eta(11)$	$k_2 = \eta(11)$	$k_3 = \eta(101)$	P	q
W_1	0	0	0	0,009079	0,6769
W_2	2	0	0	0,013824	0,4707
W_3	3	2	0	0,007935	0,7205

Отримані результати показують, що однакова кількість одиничних подій не означає однакої структури послідовності. Перехід від ізольованого розташування одиниць до їх групування змінює значення k_1 та k_2 , а відповідно — і ймовірність спільної реалізації структурних характеристик. Зокрема, конфігурація W_3 , що містить неперервну групу одиниць, характеризується меншою ймовірністю $P = 0,007935$ та більшою оцінкою структурної нетиповості $q = 0,7205$.

Однак результати не слід інтерпретувати як монотонну залежність нетиповості від ступеня групування. Наприклад, для W_1 отримано $q = 0,6769$, а для W_2 - $q = 0,4707$. Це є важливим результатом: модель оцінює **ймовірність конкретної структурної конфігурації**, а не використовує наперед задане правило «чим більше групування, тим більша аномальність».

Дослідження секвенційного агрегування. На другому етапі сформовано три сценарії зміни локальної структурної нетиповості. Для виключення впливу випадково обраних числових значень локальні оцінки взято з фактично отриманих точних розподілів для $n = 16$.

Секвенційний інтервал прийнято рівним $s = 5$, контрольний рівень — $\theta = 0,7$. Для нормування компоненти мінливості

використано $V_{\max} = 0,5$, а для компоненти напряму зміни — максимально можливий для п'ятиелементного інтервалу коефіцієнт нахилу $\beta_{\max} = 0,3$.

Розглянуто такі сценарії:

S1 — стабільний режим:

$E_1 = (0,095; 0,207; 0,095; 0,207; 0,095)$;

S2 — одиничне структурне відхилення:

$E_2 = (0,095; 0,207; 0,903; 0,207; 0,095)$;

S3 — стійка спрямована зміна:

$E_3 = (0,309; 0,498; 0,704; 0,801; 0,903)$.

У першому сценарії локальні оцінки залишаються на низькому рівні та не демонструють спрямованої зміни. У другому в середині інтервалу виникає одиничне високе значення $e_r = 0,903$, після якого система повертається до початкового режиму. У третьому сценарії локальні оцінки послідовно зростають і останні три значення перевищують контрольний рівень.

За формулами (1) – (5), наведеними в попередньому підрозділі, для кожного сценарію розраховано п'ять компонентів секвенційно агрегованої оцінки. Результати наведено в табл. 3. Результати демонструють принципово різну реакцію компонентів моделі на три сценарії. Для стабільного режиму S1 усі характеристики, крім незнач-

ної природної мінливості, мають низькі значення. Одиначне відхилення $S2$ збільшує середній рівень до $L=0,301$ і особливо мінливість до $V=0,610$, проте не формує

спрямованого тренду ($T=0$) та тривалості ($D=0$). Отже, одиначне високе значення не інтерпретується моделлю як стійка структурна зміна.

Таблиця 3.

Результати секвенційного агрегування для експериментальних сценаріїв

Сценарій	L	T	C	D	V
S1 — стабільний режим	0,140	0,000	0,000	0,000	0,109
S2 — одиначне відхилення	0,301	0,000	0,041	0,000	0,610
S3 — стійка зміна	0,643	0,497	0,061	0,600	0,428

Інша картина спостерігається для $S3$. Послідовне зростання локальних оцінок приводить до $L=0,643$, додатного напрямку зміни $T=0,497$ та накопичення $C=0,061$. Найсуттєвішою відмінністю є $D=0,600$, оскільки три з п'яти останніх оцінок утворюють неперервну серію перевищення контрольного рівня. Таким чином, одночасне зростання L , T , C і D характеризує не одиначне відхилення, а формування стійкої структурної зміни.

Особливо показовим є порівняння $S2$ і $S3$. Максимальне локальне значення в обох сценаріях практично однакове — близько 0,903. Отже, ухвалення рішення лише за максимальним значенням локальної оцінки не дозволило б розрізнити ці ситуації. Багатокомпонентне секвенційне представлення, навпаки, чітко відокремлює одиначний викид ($T=0$; $D=0$; $V=0,610$) від стійкої спрямованої зміни ($T=0,497$; $D=0,600$; $V=0,428$).

Отже, проведений експеримент підтверджує дві ключові властивості запропонованої моделі. Багатокомпонентне структурно-ймовірнісне оцінювання дозволяє розрізняти короткі послідовності, однакові за кількістю подій, але різні за характером їх локального розташування. Секвенційне агрегування, у свою чергу, дозволяє розрізняти стабільний режим, одиначне структурне відхилення та стійку спрямовану зміну за сукупністю характеристик рівня, напрямку, накопичення, тривалості та мінливості.

Практична апробація моделі для виявлення зміни стану IoT-системи

Практичну апробацію запропонованої моделі проведено на даних IoT-моніторингу прісноводної водойми. Сенсорна система забезпечувала спостереження за параметрами якості води, зокрема розчинним киснем і рН. Експериментальний масив містив 180 послідовних вимірювань; дані аналізувалися у перекривних вікнах довжини 30 спостережень із кроком 6. Початкові сигнали перетворювалися на бінарні послідовності, де «1» відповідає допустимому значенню контрольованого параметра, а «0» — його відхиленню.

Для кожного локального блока за точними спільними розподілами дво- і триланцюжків були одержані три структурно-ймовірнісні характеристики: відхилення Δ_1 , відхилення Δ_2 та спільна ймовірність $P(k_1, k_2, k_3)$. Початкові результати містять 24 послідовні блоки та демонструють перехід від стабільного режиму через формування локальних залежностей до критичного відхилення і подальшого часткового відновлення.

Для застосування розробленої в цій роботі багатокомпонентної моделі зазначені характеристики були приведені до єдиної шкали $[0;1]$: $q_{1,r} = \frac{|\Delta_{1,r}|}{\max_r |\Delta_{1,r}|}$, $q_{2,r} = \frac{|\Delta_{2,r}|}{\max_r |\Delta_{2,r}|}$, $q_{3,r} = \frac{P_r}{\max_r P_r}$.

Оскільки на етапі апробації всі три характеристики вважалися рівнозначними,

локальна інтегральна оцінка визначалася їх середнім значенням відповідно до моделі структурного агрегування, наведеної в попередньому підрозділі. Максимальні значення, використані для нормування, становили: $\max |\Delta_1| = 0,07868$, $\max |\Delta_2| = 0,13656$, $\max P = 0,18954$.

Таке перетворення не вводить нового критерію стану, а лише приводить різ-

норідні структурні характеристики до спільної шкали, необхідної для формування послідовності локальних оцінок e_r .

У табл. 4 наведено характерні результати для окремих блоків, що відповідають різним фазам розвитку досліджуваного процесу.

Таблиця 4.

Формування локальних оцінок за даними IoT-моніторингу

Блок	Δ_1	Δ_2	$P(k_1, k_2, k_3)$	q_1	q_2	q_3	e_r
1	0,00000	0,00000	0,00451	0,000	0,000	0,024	0,008
3	-0,03679	0,09589	0,01824	0,468	0,702	0,096	0,422
9	-0,07868	0,13656	0,00731	1,000	1,000	0,039	0,680
12	0,05149	0,01553	0,04404	0,654	0,114	0,232	0,333
16	-0,07868	0,13656	0,18954	1,000	1,000	1,000	1,000
20	0,00498	0,04844	0,01868	0,063	0,355	0,099	0,172
24	0,00498	0,04844	0,05199	0,063	0,355	0,274	0,231

Початкові значення Δ_1 , Δ_2 та $P(k_1, k_2, k_3)$ взято з результатів експериментального моніторингу. Для блоків 1–3 вихідне дослідження фіксує переважно стабільний режим, у блоках 4–10 з'являються локальні структурні відхилення, у блоках 11–17 залежності посилюються, а для блока 16 досягається максимальний рівень інтегрального ризику. Після цього спостерігається часткове відновлення, хоча залишкова нестабільність зберігається.

Отримана траєкторія e_r була використана як вхід другого рівня запропонованої моделі. Для практичної апробації прийнято інтервал секвенційного аналізу $s=5$ та контрольний рівень $\theta=0,6$. Для кожного положення інтервалу за раніше введеними залежностями визначено рівень L_r , напрям зміни T_r , накопичення C_r , тривалість D_r та мінливість V_r . Результати для характерних моментів наведено у табл. 5.

Таблиця 5.

Багатокомпонентні секвенційно агреговані оцінки стану IoT-системи

Кінцевий блок інтервалу	L_r	T_r	C_r	D_r	V_r	Інтерпретація
5	0,157	0,110	0,000	0,000	0,282	переважно стабільний стан
9	0,248	0,359	0,016	0,200	0,431	формування структурних відхилень
12	0,333	-0,013	0,016	0,000	0,368	підвищений, але нестійкий рівень
16	0,467	0,444	0,080	0,200	0,533	виражена спрямована зміна
20	0,388	-0,552	0,080	0,000	0,642	зниження після критичного відхилення
24	0,113	0,041	0,000	0,000	0,155	наближення до стабільного режиму

Результати показують, що багатокомпонентне представлення дозволяє описати **не лише величину відхилення, а й фазу його розвитку**. Уже до блока 9 спостерігається додатний напрям зміни $T_r=0,359$ при появі накопичення і неперервності перевищень. Для блока 16 одночасно зростають рівень $L_r=0,467$, напрям зміни $T_r=0,444$, накопичення $C_r=0,080$ і мінливість $V_r=0,533$. Саме цей блок у вихідних експериментальних даних характеризувався максимальними значеннями структурних відхилень і максимальним значенням інтегрального показника $P(k_1, k_2, k_3)=0,18954$.

Після проходження критичної ділянки характер компонентів змінюється. Для інтервалу, що завершується блоком 20, компонента напрямку набуває від'ємного значення $T_r=-0,552$, що відображає зменшення структурної нетиповості після пікового стану. Водночас підвищена мінливість $V_r=0,642$ свідчить, що стабілізація ще не є остаточною. Для блока 24 рівень знижується до $L_r=0,113$, накопичення та тривалість дорівнюють нулю, а мінливість зменшується до $V_r=0,155$, що узгоджується з поступовим відновленням, зафіксованим у вихідних даних.

Отже, практична апробація підтвердила, що запропонована модель дозволяє перетворити послідовність локальних структурно-ймовірнісних оцінок на інтерпретоване структурно-динамічне представлення стану IoT-системи. На відміну від використання одного локального показника, компоненти L_r , T_r , C_r , D_r та V_r дають змогу окремо характеризувати формування відхилення, його спрямоване посилення, критичний перехід та подальшу стабілізацію. Це підтверджує практичну придатність моделі для потокового моніторингу IoT-систем за короткими послідовностями сенсорних спостережень.

Обговорення результатів та перспективи розвитку

Отримані результати підтвердили доцільність використання багатокомпонентного підходу для аналізу коротких послідовностей. На відміну від оцінювання за

однією статистичною характеристикою, запропонована модель враховує декілька взаємопов'язаних структурних ознак і дає змогу розрізняти послідовності, що мають однакову кількість окремих подій, але відрізняються характером їх локального розташування. Експериментальне дослідження також показало, що структурна нетиповість не обов'язково монотонно зростає зі ступенем групування подій, а визначається статистичною допустимістю конкретної структурної конфігурації.

Секвенційне агрегування доповнює локальне оцінювання інформацією про характер розвитку процесу. Використання компонент рівня, напрямку зміни, накопичення, тривалості та мінливості дозволяє відокремлювати одиничні локальні відхилення від стійких спрямованих змін. Зокрема, експериментальні сценарії з однаковими максимальними значеннями локальної оцінки формували суттєво різні значення трендової компоненти, тривалості та мінливості, що підтверджує необхідність збереження порядку локальних оцінок при їх агрегуванні.

Практична апробація на даних IoT-моніторингу прісноводної екосистеми показала, що багатокомпонентна секвенційно агрегована оцінка відображає різні фази зміни стану: формування локальних відхилень, їх посилення, критичний перехід та подальше відновлення. Водночас математичне ядро моделі не залежить від фізичної природи контролюваного параметра, що створює передумови для її застосування в інших задачах потокового моніторингу технічних, інформаційних та кіберфізичних систем. Саме предметна незалежність ядра та можливість його використання для різних типів сенсорних і подієвих даних є важливою властивістю запропонованого підходу.

Разом із тим результати залежать від параметрів моделі: довжини локального вікна n , кроку його зміщення, довжини секвенційного інтервалу s , способу нормування локальних оцінок та вибору контрольного рівня θ . Тому перспективним напрямом подальших досліджень є розроблення адаптивних процедур вибору цих параметрів зале-

жно від характеристик потоку даних і типу контрольованого процесу.

Подальший розвиток моделі також доцільно пов'язати з автоматизованим визначенням ваг локальних структурно-ймовірнісних компонентів, дослідженням альтернативних операторів структурного агрегування та розширенням набору секвенційних характеристик. Окремим напрямом є використання вектора A_r як набору ознак для моделей класифікації, регресії та прогнозування, а також його інтеграція до складу інформаційних технологій раннього виявлення і прогнозування небезпечних станів. У дисертаційній моделі така траєкторія прямо розглядається як основа для виявлення моментів структурного переходу, формування правил оцінювання стану та побудови прогнозних моделей.

Висновки

У роботі розроблено багатокомпонентну модель секвенційного агрегованого оцінювання коротких послідовностей для виявлення змін стану системи. Модель реалізує два взаємопов'язані рівні перетворення: структурне агрегування локальних структурно-ймовірнісних оцінок окремого короткого фрагмента та секвенційне агрегування впорядкованої траєкторії локальних оцінок.

Основним результатом є формування багатокомпонентної секвенційно агрегованої оцінки, компоненти якої характеризують відповідно рівень, напрям зміни, накопичення, тривалість і мінливість локальної структурної нетиповості. Таке представлення дозволяє розрізняти структурно різні сценарії розвитку процесу, зокрема стабільний стан, одиничні відхилення, накопичувальні, стійкі, спрямовані та коливальні зміни. Це безпосередньо відповідає сформульованому результату наукової новизни дисертаційного дослідження.

Експериментальне дослідження підтвердило, що багатокомпонентне структурно-ймовірнісне оцінювання дозволяє розрізняти короткі послідовності, однакові за кількістю окремих подій, але різні за внутрішньою структурою. Секвенційне агрегування, у свою чергу, забезпечує відокрем-

лення випадкових локальних відхилень від стійких змін за рахунок спільного аналізу рівня, тренду, накопичення, тривалості та мінливості.

Практична апробація на даних IoT-моніторингу прісноводної екосистеми показала відповідність динаміки компонентів моделі фактичним фазам зміни стану системи – від стабільного режиму через формування і посилення структурних відхилень до критичного переходу та подальшої стабілізації. Це підтверджує можливість використання запропонованої моделі як математичної основи інформаційних технологій потокового моніторингу та раннього виявлення змін стану систем за обмеженого обсягу спостережень.

Перспективи подальших досліджень пов'язані з адаптивним налаштуванням параметрів моделі, розширенням набору секвенційних характеристик, автоматизованим визначенням ваг компонентів та використанням багатокомпонентної оцінки як вхідного простору ознак для задач класифікації і прогнозування.

References

1. Choi, K., Yi, J., Park, C., & Yoon, S. (2021). Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines. *IEEE Access*, 9, 120043–120065. <https://doi.org/10.1109/ACCESS.2021.3107975>
2. Lindemann, B., Maschler, B., Sahlab, N., & Weyrich, M. (2021). A survey on anomaly detection for technical systems using LSTM networks. *Computers in Industry*, 131, 103498. <https://doi.org/10.1016/j.compind.2021.103498>
3. Xu, J., Wu, H., Wang, J., & Long, M. (2022). Anomaly Transformer: Time series anomaly detection with association discrepancy. In *International Conference on Learning Representations (ICLR 2022)*. Retrieved from OpenReview
4. Tuli, S., Casale, G., & Jennings, N. R. (2022). TranAD: Deep transformer networks for anomaly detection in multivariate time series data. *Proceedings of the VLDB Endowment*, 15(6), 1201–1214. <https://doi.org/10.14778/3514061.3514067>
5. Zhang, C., Zhou, T., Wen, Q., & Sun, L. (2022). TFAD: A decomposition time series

- anomaly detection architecture with time-frequency analysis. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (pp. 2497–2507). ACM. <https://doi.org/10.1145/3511808.3557470>
6. Li, G., Yang, Z., Wan, H., & Li, M. (2022). Anomaly-PTG: A time series data-anomaly-detection transformer framework in multiple scenarios. *Electronics*, 11(23), 3955. <https://doi.org/10.3390/electronics11233955>
7. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2023). TimesNet: Temporal 2D-variation modeling for general time series analysis. In International Conference on Learning Representations (ICLR 2023). Retrieved from OpenReview
8. Yang, Y., Zhang, C., Zhou, T., Wen, Q., & Sun, L. (2023). DCdetector: Dual attention contrastive representation learning for time series anomaly detection. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 3033–3045). ACM. <https://doi.org/10.1145/3580305.3599295>
9. Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C. C., & Salehi, M. (2024). Deep learning for time series anomaly detection: A survey. *ACM Computing Surveys*, 57(1), Article 15. <https://doi.org/10.1145/3691338>
10. Sánchez-Ferrera, A., Calvo, B., & Lozano, J. A. (2025). A review on self-supervised learning in time series anomaly detection: Recent advances and open challenges. *ACM Computing Surveys*, 58(5), Article 119. <https://doi.org/10.1145/3770575>
11. Masol, V. I., & Popereshnyak, S. V. (2020). Checking the randomness of bits disposition in local segments of the (0, 1)-sequence. *Cybernetics and Systems Analysis*, 56(3), 513–520. <https://doi.org/10.1007/s10559-020-00267-0>
12. Masol, V. I., & Popereshnyak, S. V. (2021). Joint distribution of some statistics of random bit sequences. *Cybernetics and Systems Analysis*, 57(1), 160–167. <https://doi.org/10.1007/s10559-021-00337-x>

Дата першого надходження до видання: 05.07.2026

Внутрішня рецензія отримана: 27.07.2029

Зовнішня рецензія отримана: 29.07.2026

Дата рекомендації до друку: 10.08.2026

Дата публікації: 29.08.2026

Про автора:

Поперешняк Світлана Володимирівна,
кандидат фізико-математичних наук,
доцент
Popereshnyak Svitlana,
Ph.D. (physics & mathematics),
associate professor
<http://orcid.org/0000-0002-0531-9809>

Місце роботи автора:

Національний технічний університет
України «Київський політехнічний
інститут імені Ігоря Сікорського»
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute",
Державний університет інформаційно-
комунікаційних технологій
State University of Information and
Communication Technologies
тел. +38-098-645-54-62
E-mail: spopereshnyak@gmail.com