

УДК 389.62.1:165

# ОЦІНЮВАННЯ ЧАСОВИХ ХАРАКТЕРИСТИК ПОДІЙ ДИСКРЕТНИХ ПРОЦЕСІВ У ВІДПОВІДНОСТІ ДО КОНЦЕПЦІЇ GUM

**В. Левикін**, доктор технічних наук, професор, завідувач кафедри інформаційних управляючих систем,

**О. Чала**, кандидат економічних наук, доцент,

Харківський національний університет радіоелектроніки, м. Харків

*Запропоновано метод оцінювання часу виникнення подій у дискретних процесах з урахуванням невизначеності як розсіювання значень. Множина вимірювань часу однієї події процесу формується шляхом об'єднання моментів часу її виникнення з різних трас процесу, що входять до складу його лога. Вхідні дані для оцінювання відбираються відповідно до умов, що формуються на основі відношень переходу, нерозрізненості та обмеження. Результати оцінювання відображаються у виді середнього значення та інтервалу охоплення параметра часу. Сфера практичного застосування отриманих теоретичних результатів пов'язана з побудовою та аналізом моделей процесів у парадигмі process mining.*

*A method for estimating the time of occurrence of the events in the processes of dispersion of discrete measurement values is proposed. A number of time measurements for a single event is combined from the different traces of the process. Every trace is a part of process log. The input data for the assessment are selected in accordance with the terms and conditions, which are formed on the basis of three relations: transition, indistinguishability and restriction. Evaluation results are shown as the mean value and range of coverage of the time parameter. The practical application of the theoretical results is related to the construction and analysis of process models in the paradigm of process mining.*

**Ключові слова:** вимірювання, концепція GUM, дискретний процес, лог процесу, подія.

**Keywords:** measurement, concept GUM, discrete process, log, event.

**П**роблематика оцінювання часу виконання дискретних процесів, що представляються у виді послідовності подій, пов'язана з вирішенням завдань побудови моделей таких процесів. За побудови моделей здійснюється перехід від множинного описання процесів (множини упорядкованих подій) до логічних, алгебраїчних та інших форм представлення процесу. Відображення часу в таких моделях може бути як якісним, так і кількісним. За якісного описання часу модель дозволяє передбачити появу визначених подій у майбутньому, наприклад: у наступний момент часу; у довільний момент у майбутньому; після іншої визначеної події. За кількісного описання часу модель дозволяє визначити, наприклад, часові обмеження для подій у майбутньому, загальний час виконання процесу тощо.

Особливо актуальною проблематика визначення часових параметрів дискретних процесів є за побудови моделей методами process mining (інтелектуального аналізу процесів). Методи process mining направлені на побудову моделей дискретних процесів на основі аналізу логів, що фіксують їх виконання в рамках інформаційної системи [1]. Наявні підходи до моделювання процесів у парадигмі process mining орієнтовані значною мірою на побудову якісних моделей процесів.

Власники процесів, як правило, мають потребу в аналізі моделей, отриманих у результаті застосування методів process mining. Аналіз дозволяє виявити проблемні зони процесу і, в подальшому, реконструювати такі фрагменти для підвищення їх ефективності. У проблемних зонах звичайно виникають затримки під час виконання



В. Левикін



О. Чала

процесу. Наприклад, типові проблеми процесного управління для одного із підрозділів фірми Volvo пов'язані з перекладанням відповідальності між виконавцями процесу та зі зловживанням статусом «в очікуванні користувача»[2]. Очевидно, що перша проблема призводить до затримок внаслідок необґрунтованих змін виконавців операцій процесу, а друга вказує на спробу приховати затримки в обслуговуванні користувачів.

Наведені проблеми важко виявити в силу багатоваріантності процесів: за кожної реалізації процес адаптуватиме час виникнення подій у відповідності з умовами зовнішнього середовища. Це призводить до відповідних змін у лозі процесу.

Традиційно, за побудови моделі процесу виконується об'єднання подій з урахуванням їх часового параметра [3]. У результаті отримуємо модель процесу з розгалуженням часу, що відображає його можливі траси. За такого підходу в одній моделі об'єднуються дії для різних рівнів ієрархії об'єкта, на якому проходить розглянутий процес. На практиці, за використання логів діючих підприємств, отримуємо так звану «спагеті» модель [4]. Аналіз і, зокрема, оцінювання часу виконання процесів для такої моделі пов'язані з труднощами і потребують декларативних підходів до побудови моделей процесу [5, 6].

Отже, сучасні методи інтелектуального аналізу процесів направлені на побудову моделей, що відображають всі можливі варіанти реалізації процесу. Разом з тим практичні потреби аналізу процесів пов'язані з виявленням «вузьких місць» на основі порівняння як моделей, так й їх логів. «Вузькі місця» — це фрагменти процесу, в яких відбуваються затримки часу за його проходження. Тому для порівняння процесів та виявлення фрагментів з втратами часу потрібно визначити такий спосіб оцінювання часу подій для логів процесів, який би дозволив порівнювати ефективність самих процесів у часовому вимірі.

Зазначене вище свідчить стосовно актуальності теми дослідження.

*Мета цього дослідження* — розроблення методу оцінювання часу виникнення подій (що фіксують виконання відповідних дій) дискретних процесів з урахуванням невизначеності як розсіювання значень у відповідності з концепцією GUM.

У практичному плані оцінка часу виконання процесів повинна враховувати як затримку, так і більш раннє виникнення окремих подій процесу для того, щоб можна було порівняти аналогічні процеси на різних підприємствах та виявити «вузькі місця».

## ОЦІНЮВАННЯ ЧАСОВИХ ХАРАКТЕРИСТИК ПОДІЙ ДИСКРЕТНИХ ПРОЦЕСІВ

У відповідності з концепцією GUM для побудови моделі вимірювань потрібно виконати такі кроки на етапах формулювання та обчислень[7]:

На етапі формулювання необхідно: визначити вихідні (ті, що вимірюються) та вхідні величини; розробити модель вимірювань; визначити розподіл ймовірностей для вхідних величин.

На етапі обчислень у загальному випадку виконується трансформація розподілу ймовірностей для вхідних величин за допомогою моделі вимірювань для того, щоб отримати розподіл вихідної величини.

Для визначення вхідних величин розглянемо особливості вхідних логів, що використовуються для побудови моделей дискретних процесів методами інтелектуального аналізу процесів.

Традиційний підхід до побудови моделі процесу в парадигмі process mining передбачає об'єднання послідовностей подій процесу. Лог містить множину послідовностей подій, кожна з яких відображає одноразове виконання процесу. Таку послідовність подій звичайно визначають як слід або трасу процесу. Для подій зазначається час їх виникнення. Загалом лог об'єднує всі варіанти виконання процесу, які були використані на практиці. Відзначимо, що процеси можуть бути недокументованими. У такому випадку лог фіксує порядок вирішення поточних задач. Для документованих процесів засоби process mining дозволяють виявити відхилення описання та реального процесу, а для недокументованих — побудувати модель процесу.

Кожна подія логу фіксує виконання тієї чи іншої дії процесу. Часовий параметр кожної події — результат одноразового вимірювання часу закінчення (інколи — початку) дії процесу. Лог містить множину вимірювань часу виникнення кожної події процесу.

З точки зору технічної процедури отримання вимірювальної інформації маємо розсіювання значень, що приписують вимірюваній величині. Таке розсіювання, у відповідності з концепцією GUM, характеризується терміном «невизначеність» [8, 9] і відповідає залишковій невизначеності знань стосовно вимірюваної величини [10, 11]. Тому за формування інтегрального показника оцінювання часу виконання процесів, що охоплює час виконання вибраних дій процесу для всіх його наявних трас, доцільно використовувати концепцію GUM.

Оскільки за вимірювання часу потрібно враховувати розсіювання для різних подій процесу, ці події повинні відобразитися в одному масштабі часу.

Відзначимо, що побудова моделі процесу методами process mining базується на припущенні, що для всіх подій використовується однакова деталізація часу. Проте ця умова виконується не завжди, оскільки масштаб часу значною мірою залежить від властивостей об'єктів, що беруть участь у процесі (наприклад, елементів комп'ютерної системи, інтелектуальних агентів, людей-виконавців). Одна траса процесу може охоплювати події, які потрібно аналізувати у різному масштабі часу [12].

Тому для оцінювання часу виникнення подій у таких процесах необхідно розподілити події за різними рівнями деталізації часу.

Багаторівнева деталізація часу, запропонована авторами, базується на загальних властивостях грануляції [13], а також підходах до грануляції часу [14,15].

Така деталізація часу містить такі складові:

1) Множину релевантних наборів моментів часу. Кожний із наборів відповідає множині подій з одним рівнем деталізації часу. Елементи одного набору пов'язані відношеннями переходу та порядку.

2) Підмножини подій, що належать до одного рівня деталізації та пов'язані відношенням нерозрізненості.

3) Підмножини подій одного рівня деталізації, між якими встановлено відношення обмеження. Це відношення дозволяє розділити окремі події або їх групи.

4) Підмножини подій, між якими встановлено відношення узагальнення/деталізації. Такі підмножини належать до різних рівнів деталізації часу.

У відповідності з цим підходом за оцінювання часу виникнення подій використовують складові 1—3. Розглянемо їх детальніше.

Релевантний набір підмножин моментів часу  $T_i \subseteq T$  описує послідовність подій (у цій моделі моментів часу  $t_{ij} \in T_i$ , коли відбувалися події процесу) на різних рівнях деталізації часу. Між подіями з кожної підмножини  $T_i$  на одному рівні деталізації задані відношення переходу та порядку. Відношення переходу — антирефлексивне. Традиційно для нього виконується умова:

$$t' < t'' \Rightarrow \neg \exists t : t' < t \wedge t < t'', t', t'', t \in T_i. \quad (1)$$

У відповідності з (1) для пари моментів  $t'$  і  $t''$ , між якими встановлено відношення переходу, неможливо досягти більшої деталізації часу, оскільки для підмножини  $T_i \subseteq T$  процесу не існує проміжних моментів  $t$  між  $t'$  і  $t''$ . Зазначене співвідношення — істинне для одного рівня деталізації процесу. За традиційного підходу до побудови моделі на основі об'єднання трас співвідношення (1) може не виконуватися у таких випадках:

- відбувається злиття паралельних послідовностей подій;

- маємо розсіювання моментів часу для однієї й тієї ж події на різних трасах процесу.

У першому випадку доцільно розмістити набори моментів часу, що відповідають паралельним послідовностям дій, на окремих рівнях деталізації часу.

У другому — за визначення відношення переходу необхідно враховувати непевність у вимірюванні часу подій для різних трас. Визначимо через  $\tau_j$  підмножину відносних моментів часу  $t_{kj} \in \tau_j$ , що відповідають виникненню  $j$  події на різних трасах процесу. Тобто  $\tau_j$  містить  $k$  вимірювань моментів часу виникнення  $j$  події.

Тоді для кожного елемента із підмножини  $\tau_j$  виконується відношення нерозрізненості (2):

$$t' \approx t'' \equiv ((t' < t'') \vee (t'' < t')) \vee (t' < t') \vee (t'' < t'') = \text{false} \forall t', t'' \in \tau_j. \quad (2)$$

Відношення порядку задається через відношення переходу:

$$t' < t'' \Rightarrow \exists t : t' < t \dots < t''. \quad (3)$$

Відповідно, кожна множина  $\tau_j$  містить лише ті моменти часу, для яких неможливо задати відношення переходу:

$$\forall t' \in \tau_j, \exists t'' \in \tau_j : t' \approx t''. \quad (4)$$

Нерозрізненість елементів підмножини  $\tau_j$  у рамках запропонованого підходу свідчить стосовно непевності у вимірюванні часу виникнення  $j$  події згідно з концепцією GUM. Тому необхідно визначити відношення переходу для  $j$  і  $j+1$  подій через відношення між підмножинами  $\tau_j$  і  $\tau_{j+1}$ :

$$\tau_j < \tau_{j+1} \text{ iff } \exists t' \in \tau_j, t'' \in \tau_{j+1} : t' < t''. \quad (5)$$

Із визначення (5) видно, що умова для встановлення відношення переходу для підмножин  $\tau_j$  і  $\tau_{j+1}$  — наявність хоча б однієї пари елементів із цих підмножин, для яких виконується відношення переходу. Тому наявність відношення (5) для процесу визначається наявністю відношень  $t' < t''$  у вхідному лозі.

Очевидно, що за визначення переходу в сенсі (5) можуть існувати пари елементів  $t^* \in \tau_j, t^{**} \in \tau_{j+1}$  такі, що  $t^{**} < t^*$ . Тобто моменту настання другої події може передувати момент настання першої на шкалі часу. Це означає порушення лінійності часу в моделі й може свідчити стосовно неефективності ділянки процесу  $(\tau_j, \tau_{j+1})$ . Тому необхідна умова для аналізу часових характеристик процесу ділянки  $(\tau_j, \tau_{j+1})$ :

$$\exists t^* \in \tau_j, t^{**} \in \tau_{j+1} : t^{**} < t^* \mid \tau_j < \tau_{j+1}. \quad (6)$$

Умова (6) дозволяє визначити відношення обмеження  $\tau_j \oplus \tau_{j+1}$ . Останнє використовують для того,

щоб розділити події підмножин  $\tau_j$  і  $\tau_{j+1}$ :

$$\tau_j \oplus \tau_{j+1} \text{ iff } (\tau_j \prec \tau_{j+1}) \wedge (t' \prec t'') \quad (7)$$

$$\forall t' \in \tau_j, t'' \in \tau_{j+1}$$

Нерозрізнені елементи  $\tau_j$  і  $\tau_{j+1}$  відображають непевність вимірювань часу виникнення відповідних подій. Тому з підмножин  $\tau_j$  і  $\tau_{j+1}$  доцільно вилучити ті елементи, для яких не виконується умова (7). У практичному аспекті це означає, що як вхідні дані не враховують ті моменти часу, для яких порушується умова лінійності часу. Таке порушення свідчить стосовно затримки в ході реалізації процесу.

Отже, вхідні дані для оцінювання часу визначаються умовами (1)–(3) для окремих моментів часу та умовами (4), (5), (7) для підмножин  $\tau_j$  і  $\tau_{j+1}$ , що містять дані вимірювань часу  $t$  для  $j$  та  $j+1$  подій відповідно.

За виконання наведених умов кожне вимірювання часу  $t$  для  $j$  події належить до інтервалу:

$$t \in [t_{j\min}, t_{j\max}] \quad \forall t \in \tau_j \quad (8)$$

$$(t' < t_{j\min}) \wedge (t'' > t_{j\min}) \quad \forall t' \in \tau_{j-1}, t'' \in \tau_{j+1}$$

Із (8) видно, що наведені вище відношення задають обмеження для вхідних даних оцінювання часу виникнення подій дискретних процесів. Ці обмеження дозволяють виключити із вхідних даних ті моменти часу, які характеризують неефективне виконання процесу.

У рамках задачі дослідження нас цікавить належність вхідних даних до інтервалу (8), а не розподіл цих даних. Тому, відповідно до [16], знання щодо вхідної величини характеризується прямокутним розподілом з межами  $t_{j\min}$  і  $t_{j\max}$ .

Модель вимірювань у відповідності з наведеним підходом дає можливість отримати оцінку часу виконання кожної  $j$  події та інтервал охоплення.

Значення часу для підмножини  $\tau_j$  визначаємо як середнє значення моментів часу:

$$t_{\text{ср}} = \frac{\sum t_{kj}}{K} \mid t_{kj} \in [t_{j\min}, t_{j\max}], k = \overline{1, K} \quad (9)$$

Інтервал охоплення, що містить сукупність істинних значень вимірюваної величини із заданою вірогідністю, визначається відповідно до виразу (8).

Інтервал охоплення дозволяє визначити можливість оптимізації процесів на основі співставлення мінімальних  $[t_{j-1\max}, t_{j\min}]$  та максимальних  $[t_{j-1\min}, t_{j\max}]$  інтервалів часу між двома послідовними подіями  $j$  та  $j+1$ .

Середні значення часу для підмножини  $\tau_j$  дають можливість прогнозувати час виконання процесу.

Умова (6) дозволяє відділити події, що не входять

до інтервалу охоплення, і, отже, визначити найменш ефективні в темпоральному сенсі фрагменти процесу. Ця умова забезпечує можливість для виявлення трас логу з несвоечасним виникненням попередньої  $j-1$  події, що потенційно може затримати наступну  $j$  подію процесу.

Метод оцінювання часових характеристик подій дискретних процесів, що базується на наведених обмеженнях і моделі, використовує як вхідні дані набір підмножин моментів часу, для яких задані відношення переходу.

Метод містить такі етапи.

1. Нормування часу. На цьому етапі виконується перехід від абсолютних до відносних значень часу, що дозволяє порівняти послідовності подій, які належали до різних трас процесу і тому відбувалися в різні проміжки часу.

2. Визначення підмножин  $\tau_j$ , для яких виконується умова нерозрізненості (4). На цьому етапі формується початковий набір підмножин вхідних даних шляхом об'єднання результатів фіксації часу для кожної  $j$  події в різних трасах процесу.

3. Визначення часових обмежень (7) для пар подій  $j$  та  $j+1$ , для яких задано відношення переходу. На цьому етапі формується інтервал охоплення часу для кожної події, а також множина трас, що потребують подальшого аналізу позицій ефективності в часовому вимірі.

4. Оцінювання значення часу для кожної події відповідно до (9).

У рамках цього дослідження розглянуто лише відношення переходу для дискретного процесу. Однак наведені вище кроки з формування набору вхідних даних для оцінювання часу подій також можна застосовувати для подій, пов'язаних відношенням порядку (3). Відношення порядку зводиться до відношення переходу шляхом вилучення проміжних подій.

## ВИСНОВКИ

Запропоновано модель та метод оцінювання часових характеристик подій дискретних процесів із використанням концепції вимірювань GUM.

Модель відрізняється тим, що враховує невизначеність як розсіювання вимірювань часу для однакових подій з різних реалізацій одного дискретного процесу. Множина вимірювань часу однієї події процесу формується шляхом об'єднання моментів часу її виникнення з різних трас процесу, що входять до складу його логу. Вхідні дані моделі вимірювань відбираються відповідно до обмежень на основі відношень переходу, нерозрізненості та обмеження.

Модель дозволяє визначити середнє значення та інтервал охоплення параметра часу.

Метод базується на моделі оцінювання і передбачає нормування часу подій, формування підмножин моментів виникнення кожної події, фільтрацію тих моментів виникнення подій, що не відповідають нормальному виконанню процесу, а також отримання оцінки часу виникнення та інтервалу охоплення часу кожної події.

У практичному плані метод створює умови для порівняння розподілених у просторі аналогічних процесів або процесів із подібних предметних сфер. Наприклад, з'являється можливість порівняти темпоральні характеристики однакових процесів на різних підприємствах із урахуванням розсіювання подій. Такий порівняльний аналіз дозволяє, зокрема, виявити умови виникнення «проблемних» подій, які призводять до затримок у виконанні процесу.

### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ / REFERENCES

1. W.M.P. van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes [Text] / W.M.P. van der Aalst. Springer-Verlag, Berlin, 2011. — 352 p.
2. Dongen, van B.F. Business Process Intelligence Challenge // B.F. van Dongen, B. Weber, D.R. Ferreira, and J. De Weerd. [Электронный ресурс] / Режим доступа: www/ URL: <http://ceur-ws.org/Vol-1052/summary.pdf/> — 2013 г.
3. W.M.P. van der Aalst. Process Mining in the Large: A Tutorial. [Text] / In E. Zimnyi, editor, Business Intelligence (eBISS 2013), volume 172 of Lecture Notes in Business Information Processing, Springer-Verlag, Berlin, 2014. — P. 33—76.
4. W.M.P. van der Aalst. Process Mining: Discovering and Improving Spaghetti and Lasagna Processes [Text] / W.M.P. van der Aalst // IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2011). — Paris, France, 2011. — P. 13—20.
5. W.M.P. van der Aalst Declarative Workflows: Balancing Between Flexibility and Support [Text] / W.M.P. van der Aalst, M. Pesic, H. Schonenberg // Computer Science — Research and Development, 23(2). — 2009. — P. 99—113.
6. De Leoni M. Aligning event logs and Declarative process models for conformance checking [Text] / M. De Leoni, F. Maggi, W.M.P. vander Aalst, // Proceedings of the 10th International Conference on Business Process Management, Lecture Notes in Computer Science, vol. 7481, Springer, Berlin, Heidelberg, 2012. — P. 82—97.
7. Введение к «Руководству по выражению неопределенности измерения» и сопутствующим документам. Оценивание данных измерений [Text] / Пер. с англ. под науч. ред. д.т.н., проф. В.А. Слаева, д.т.н. А.Г. Чуновкиной. — СПб.: «Профессионал», 2011. — 58 с/р.
8. ISO/IEC Guide 98. Part 3: Guide to the expression of uncertainty in measurement (GUM:1995) (Частина 3: Настанови з подання непевності у вимірюванні).
9. International vocabulary of basic and general terms in metrology (VIM). (Міжнародний словник основних і загальних термінів у метрології). ISO, 1993.
10. Чалий В. Щодо обґрунтування терміна «непевність» у вимірюванні / В. Чалий, О. Чала // Метрологія та прилади. — 2011. — №4. — С/Р. 3—10.
11. Чалий В. Деякі філософсько-інтерпретаційні та прикладні аспекти концепції непевності у вимірюванні / В. Чалий // Метрологія та прилади. — 2015. — № 4(54). — С/Р. 46—49.
12. Hobbs J.R., Granularity [Text] / J.R. Hobbs // Ninth International Joint Conference on Artificial Intelligence, Los Angeles, California, 1985. — 432—435 pp.
13. C. Bettini, X. S. Wang, S. Jajodia, and J-L. Lin. Discovering Frequent Event Patterns with Multiple Granularities in Time Sequences. IEEE Transactions on Knowledge and Data Engineering, 2(10):222—237, 1998.
14. Bettini C. Time Granularities in Databases, Data Mining, and Temporal Reasoning. [Text] / C. Bettini, S. Jajodia, and X.S. Wang // Springer, Berlin, Germany, 2000. — 230p.
15. Montanari A.. Metric and Layered Temporal Logic for Time Granularity. [Text] / A. Montanari. — PhD thesis, University of Amsterdam, Amsterdam, Netherlands, september 1996. — 178p.
16. ISO/IEC Guide 98-3:2008, Неопределенность измерения — Часть 3: Руководство по выражению неопределенности измерения (GUM:1995) JCGM 100:2008

Отримано / received: 01.12.2015.

Стаття рекомендована до публікації д.т.н., проф. В.О. Філатовим (Україна).  
Prof. V.O. Filatov, D. Sc. (Techn.), Ukraine, recommended this article to be published.