

УДК 004.9: 519.237.8

О. Г. Байбуз, М. Г. Сидорова, Т. В. Вершиніна

Дніпровський національний університет імені Олеся Гончара

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ДИНАМІКИ ШВИДКОСТІ РЕАКЦІЇ ОПЕРАТОРІВ

Розроблено інформаційну технологію, яка дозволяє визначати схожі закономірності динаміки швидкості реакції різних психофізіологічних типів особистостей, класифікувати нові об'єкти спостереження до найбільш відповідного психотипу, а також аналізувати зміни динаміки в поведінці кожного з досліджуваних операторів.

Ключові слова: інформаційна технологія, часові ряди, міри подібності, швидкість реакції, сегментація багатовимірних часових рядів.

Разработана информационная технология, которая позволяет определять похожие закономерности динамики реакции различных психофизиологических типов личностей, классифицировать новые объекты наблюдения к наиболее подходящему психотипу, а также анализировать изменения динамики в поведении каждого из исследуемых операторов.

Ключевые слова: информационная технология, временные ряды, меры сходства, скорость реакции, сегментация многомерных временных рядов.

The purpose of this work is to develop the information technology for studying the dynamics of adaptive properties of the nervous system of operators. Experimental data are multi-dimensional time series that have been obtained using the reflexometric method of analyzing the reaction time of various psychophysiological types of individuals. For efficient processing of these data, better understanding the structure, finding hidden patterns, trends, some algorithms are proposed, and software (Java programming language) that implements them and provides visualization of the data and the obtained results for convenient analysis is created.

The proposed technology allows to determine the similar patterns of the dynamics of reaction rates of various psychophysiological types of individuals, to classify new observation objects to the most relevant psychotype, and to analyze changes of the dynamics in the behavior of each of the investigated operators.

The detection of the similarities of the dynamics series, for a more objective assessment, is carried out with the application of a set of similarity measures of different categories: Shape-, Edit-, Structure-, Model- and Compression-based (DTW, LCSS, modified LCSS, TQuest, Euclidean metrics, correlation metrics, trend, seasonality indicators, etc., construction of dendrograms by hierarchical clusterization method). To classify a new object of observation to the most

appropriate psychotype, a new method based on a multicriteria comparison with the models is proposed. Analysis of changes in the dynamics of the reaction rate of each of the investigated operators is based on the segmentation algorithm of the multidimensional series on the basis of differential evolution with the previous smoothing of data.

The proposed information technology has been tested on the data of the reaction time dynamics of various psychophysiological types of individuals, but can also be applied in the analysis of the similarity in the dynamics of processes of different nature.

Keywords: information technology, time series, similarity measures, reaction rate, segmentation of multidimensional time series.

Вступ. Дослідження адаптаційних властивостей нервової системи має велике значення в теоретичних і прикладних дослідженнях. Існує багато методів та експрес-методик діагностики основних властивостей нервової системи. Важливим серед рефлексометричних методів є аналіз часу реакції.

Час реакції – основний поведінковий параметр в експериментальній психофізіології. Використовується для оцінки швидкості реакції оператора. Враховується при проектуванні та розробці найрізноманітніших технічних пристроїв, є критерієм придатності до багатьох професій, таких як оператор, диспетчер, шофер та ін.

Для ефективної обробки експериментальних даних, аналізу прихованих у них закономірностей, тенденцій, кращого розуміння структури актуальною задачею є розробка інформаційних технологій та сучасного програмного забезпечення.

Аналіз досліджень та публікацій. Для визначення схожості часових рядів використовують міри подібності [1–3]. Вони поділяються на чотири категорії: Shape-, Edit-, Structure-, Model- та Compression-based. Shape-based метрики порівнюють форму рядів в цілому. Евклідова відстань та інші L_p норми є найбільш широко використовуваними представниками цієї категорії. Проте у багатьох випадках вони є неефективними. Навіть якщо проблеми масштабування і шуму можуть бути вирішені на етапі попередньої обробки, то для врахування викривлень та викидів необхідно застосовувати більш складні методи, наприклад, метод динамічної трансформації шкали часу (Dynamic Time Warping) та його різновиди. Також до цього класу можна віднести DISSIM, Spatial Assembling Distance (SpADe) та elastic метрики. Основна ідея Edit Based підходу полягає в тому, що відстань між рядами може бути представлена як мінімальна кількість операцій, необхідних для перетворення одного ряду в інший, за допомогою операцій вставки, видалення і заміни.

Представниками є такі метрики: LCSS, FTSE, EDR, ERP, TWED та інші. Feature-based метрики засновані на обчисленні набору характеристик, що відображають різні аспекти часового ряду, є корисними для коротких серій або підпоследовностей. Structure-based підхід на відміну від Feature-based визначає структури більш високого рівня і порівнює їх у більш глобальному сенсі. Поділяється на дві підкатегорії: Model-based та Compression-based. Вибір адекватної міри подібності залежить від характеру даних, конкретних властивостей, притаманних рядам та необхідних у дослідженні.

Сегментація є важливим напрямом обробки й аналізу часових рядів [4–5]. Оскільки тенденція у довгих часових рядах може час від часу змінюватися та набувати різного характеру, потрібно виявити відповідні сегменти та точки розладнання. Хоча проблемі сегментації приділено багато уваги, вона не була вирішена з високою точністю. У літературі є три основні класи класичних алгоритмів сегментації. Top-down підхід («зверху вниз») полягає в рекурсивному поділі часового ряду, доки не буде виконано якийсь критерій зупинки. Ідея підходу близька до дивізійної ієрархічної кластеризації. Bottom-up – починаючи з деякого кращого можливого наближення, сегменти об'єднуються, доки не будуть виконані деякі критерії зупинки. Ідея підходу близька до агломеративної ієрархічної кластеризації. Sliding-window – сегмент збільшується до тих пір, поки не буде перевищено поріг помилки. Процес повторюється з наступною точкою даних, що не входить до вже визначеного сегменту. Визначення сегментів за допомогою лінійної апроксимації та еволюційних методів є поширеним у літературі.

Більшість підходів до сегментації часових рядів можуть бути застосовані лише для часових рядів однієї змінної. У випадку багатовимірних рядів сегментація є більш складною, оскільки кожен функцію можна розглядати як окремий ряд, і тому те, що можна вважати кордоном для одного ряду, не може розглядатися як межа в іншому, тобто переходи в кожній функції можуть розпочинатися й закінчуватися в різні точки часу, забезпечуючи суперечливі дані про те, де починається і закінчується сегмент. Лише деякі роботи, присвячені сегментації багатовимірних часових рядів, можна знайти в літературі, наприклад [6–8].

Постановка задачі. Об'єктом дослідження є аналіз динаміки адаптаційних властивостей людини-оператора на основі даних тесту на час реакції. Дані представлені у вигляді часових рядів $\{x_i(t), y_i(t); i = \overline{1, N}, t = \overline{1, T}\}$. Оператори (користувачі ПК), що приймали

участь у дослідженні, відносяться до різних типів особистостей за психофізіологічними показниками. Протягом п'яти хвилин, через кожні 3000 мс, у кожен з t моментів часу на екрані з'являється зображення-подразник, на який об'єкти дослідження повинні реагувати натисненням клавіші. Реєструвався час отримання відповідної реакції у вигляді натискання клавіші на появу сигналу $x_i(t)$, а також час утримання клавіші натиснутою при цьому $y_i(t)$.

У роботі поставлено за мету запропонувати інформаційну технологію для аналізу цих даних, а саме: виявлення схожих закономірностей реакції різних психофізіологічних типів особистостей, класифікації нових об'єктів спостереження до найбільш відповідного психотипу, а також аналізу зміни динаміки в поведінці кожного з досліджуваних операторів.

Основний матеріал. Для реалізації поставленої задачі розроблено інформаційну технологію, що складається з наступних етапів: порівняльний аналіз динаміки швидкості реакції операторів різних психотипів; визначення близького психотипу на основі порівняння з еталонами; сегментація часових рядів для аналізу змін динаміки швидкості реакції кожного з досліджуваних операторів. Розглянемо кожен задачу більш детально.

Порівняльний аналіз динаміки швидкості реакції операторів різних психотипів. Позначимо порівнювану пару рядів динаміки $Q = q_1, q_2, \dots, q_n$ та $C = c_1, c_2, \dots, c_n$; щоб їх порівняти, необхідно обчислити міру відстань d , що має бути мінімальною, або міру подібності s , що має бути максимальною в разі схожості рядів. Оскільки існує велика кількість мір близькості та методів порівняння часових рядів, що мають свої переваги та недоліки, в роботі пропонується для більш об'єктивної оцінки проводити аналіз за набором різних критеріїв: динамічної трансформації шкали часу (DTW); пошуку найбільшої спільної підпоследовності (LCSS); модифікації методу пошуку найбільшої спільної підпоследовності, що дозволяє контролювати зсув даних по часовій осі; методу, що враховує порогове значення (TQuest); метрик Евклідової та на основі коефіцієнта кореляції Пірсона; мір близькості (середньої, середньої квадратичної та пікової), коефіцієнтів тренда та сезонності. Далі наведено обчислювальні схеми методів DTW та LCSS.

Алгоритм DTW.

1. Будуємо матрицю відстаней d порядку $n \times m$ між двома рядами Q та C , використовуючи евклідову метрику, де

$$d_{ij} = d(q_i, c_j).$$

2. На основі матриці відстаней будуємо матрицю трансформації (деформації) D , де $D_{ij} = d_{ij} + \min(D_{i-1,j}, D_{i-1,j-1}, D_{i,j-1})$.
3. Визначаємо оптимальний шлях трансформації (деформації) W – набір суміжних елементів матриці, який встановлює відповідність між двома рядами Q та C . Він являє собою шлях, який мінімізує загальну відстань між Q та C , де k -ий елемент шляху $W = w_1, \dots, w_k, \dots, w_K$, $\max(m, n) \leq K \leq m + n$ визначається як $w_k = (i, j)_k$, $d(w_k) = d(q_i, c_j)$, де K – довжина шляху.
4. Визначаємо DTW-відстань на основі оптимального шляху трансформації за такою формулою:

$$DTW(Q, C) = \min \left\{ \frac{\sum_{k=1}^K d(w_k)}{K} \right\}$$

Алгоритм LCSS1.

1. Будуємо матрицю підпоследовностей LCSS(i, j):

$$LCSS(i, j) = \begin{cases} 0, & i = 0 \\ 0, & j = 0 \\ 1 + LCSS[i-1, j-1], & |q_i - c_j| < \varepsilon \\ \max(LCSS[i-1, j], LCSS[i, j-1]) & \end{cases}$$

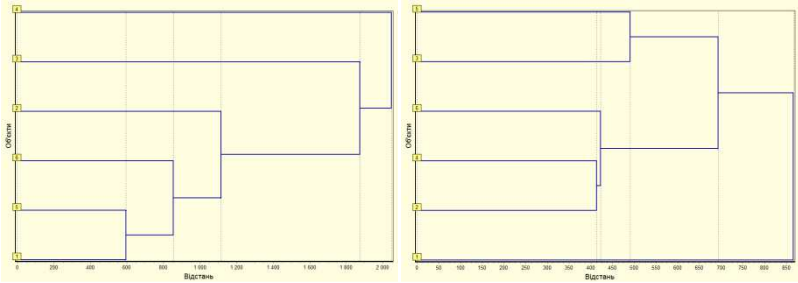
де i – індекс елементу першого ряду, а j – індекс елементу другого ряду.

2. Визначаємо найдовшу підпоследовність, яка має відповідати останньому елементу отриманої матриці.
3. Визначаємо LCSS-відстань: $LCSS = 1 - \frac{LCSS_{\varepsilon}(Q, C)}{\min(n, m)}$

Складність алгоритму $O(m \cdot n)$ можна поліпшити до $O(m+n)\delta$, якщо включити до алгоритму додаткове обмеження при обчисленні $|k-l| \leq \delta$, де k, l – довжини відповідних підпоследовностей, тобто

$$LCSS(i, j) = \begin{cases} 0, & i = 0 \\ 0, & j = 0 \\ 1 + LCSS[i-1, j-1], & |q_i - c_j| < \varepsilon, |k-l| \leq \delta \\ \max(LCSS[i-1, j], LCSS[i, j-1]) & \end{cases}$$

Для більш наочного аналізу подібності часових рядів було побудовано дендрограму методом ієрархічної кластеризації [9] (рис. 1).



а

б

Рисунок 1 – Дендрограма:

а – ряди динаміки часу натиснення на клавішу; б – ряди динаміки часу утримання клавіші натиснутою

Визначення близького психотипу на основі багатокритеріального порівняння з еталонами. Для віднесення нового об'єкта спостереження до найбільш відповідного психотипу слід порівняти його ряд динаміки з наявними еталонами психотипів за всіма запропонованими мірами подібності та обрати результат, що переважає за більшістю з них.

Аналіз змін динаміки швидкості реакції кожного з досліджуваних операторів. При роботі з рядами динаміки наявність шуму часто ускладнює аналіз структури ряду. Тому як попередній етап сегментації (виділення періодів з різною динамікою) було реалізовано згладжування даних.

При дослідженні багатовимірних рядів цікаво аналізувати одразу всі його виміри та знаходити такі точки розладнання, що будуть оптимальними та змістовними для кожного з одновимірних рядів. Ця задача є досить складною та не завжди має розв'язок, оскільки те, що можна вважати межею для одного ряду, не може розглядатися як межа в іншому, тобто зміни динаміки кожного одновимірного ряду можуть починатися і закінчуватися в різні точки часу, забезпечуючи суперечливі дані про те, де починається і закінчується сегмент. У розробленій авторами технології пропонується застосовувати підхід на основі диференціальної еволюції [7].

Розглянемо багатовимірні часові ряди вигляду $x_{t,k}, t = \overline{1, T}, k = \overline{1, M}$, де k – вимірність, припускаючи, що кількість сегментів відома апріорно, а межі сегментів – вільні параметри $[b_1, b_2, \dots, b_{c-1}]$, де $b_0 = 1, b_c = T$.

Популяція векторів задана набором $X = \{x_1, x_2, \dots, x_p\}$, де p – розмір популяції, кожний вектор x_i – складається з дійсних чисел та має розмірність $c-1$. Значення векторів популяції обираються з одиничного інтервалу $[0,1]$.

Алгоритм сегментації багатовимірного ряду на основі диференціальної еволюції:

1. Створити популяцію розміром P .
2. Поки значення критерію не задовільнено для кожного елемента популяції x_i , виконати:
 - 1) випадковим чином обрати три батьківські вектори (усі різні): $x_{r_1}, x_{r_2}, x_{r_3}$;
 - 2) згенерувати мутований вектор наступним чином: $v_i = x_{r_1} + G[x_{r_2} - x_{r_3}]$;
 - 3) створити новий вектор шляхом комбінування значень мутованого вектора та початкового: $u_i = \text{crossover}(v_i, x_i)$.
 - 4) якщо визначена фітнес-функція $f(u_i) < f(x_i)$, то замінити у вхідній популяції x_i на u_i .

Параметр G обирається довільно в діапазоні $[0;2]$. Метод дозволяє використовувати будь-які варіації операції кросоверу. Її суть полягає у комбінації двох векторів згідно з випадковою вірогідністю. Тобто, якщо вірогідність мутації не перевищує встановлений показник, то новому вектору присвоюється значення мутованого вектора, у всіх інших випадках – початкового вектора.

Розрахунок фітнес-функції. Межі сегментів визначаються через диференційну еволюцію. Показник, що характеризує підпоследовність обчислюється для кожної змінної $k = \overline{1, M}$: $D_{ik} = x_{t+1,k} - x_{t,k}$, $t = \overline{1, T-1}$.

Враховуючи набір меж $[b_1, b_2, \dots, b_{c-1}]$, визначають довжину кожного сегмента: $L_i = b_i - b_{i-1}$. Обчислюють дисперсію за допомогою виразу:

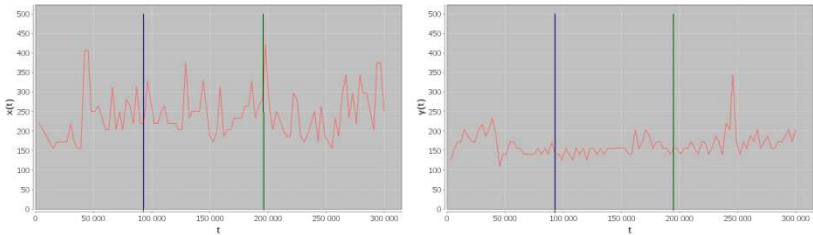
$$S_{ik}^2 = \sum_{t=b_{i-1}}^{b_i} \frac{(D_{tk} - \overline{D_{ik}})^2}{L_i - 1}, \text{ де } \overline{D_{ik}} = \frac{1}{L_i} \sum_{t=b_{i-1}}^{b_i} D_{tk}.$$

Далі розраховують S для кожного M -мірного сегменту: $S_i = \frac{1}{M} \sum_{k=0}^M \sqrt{S_{ik}^2}$.

Сегментація виконується шляхом мінімізації наступної фітнес-

функції: $f = \sum_{i=1}^C S_i L_i$.

На рис. 2 наведено результати сегментації двовимірного часового ряду динаміки швидкості реакції, що характеризує 6-й психотип. Отримані результати не завжди можна легко інтерпретувати, оскільки зміна характеру підпоследовностей може бути не досить наочною.



а

б

Рисунок 2 – Результати сегментації

а – динаміка часу натиснення на клавішу; б – динаміка часу утримання клавіші натиснутою

Зміну тенденції швидкості реакції можна обумовити, наприклад, втомою або навпаки запалом енергії, що може збільшуватися або зменшуватися. Аналіз, розуміння причин виникнення та прогнозування тривалості таких тенденцій є важливою задачею при аналізі поведінки операторів.

Розроблене програмне забезпечення написано мовою Java у середовищі IntelliJ IDEA 2016.1 Community Edition. Приклад інтерфейсу взаємодії з користувачем наведено на рисунку 3.



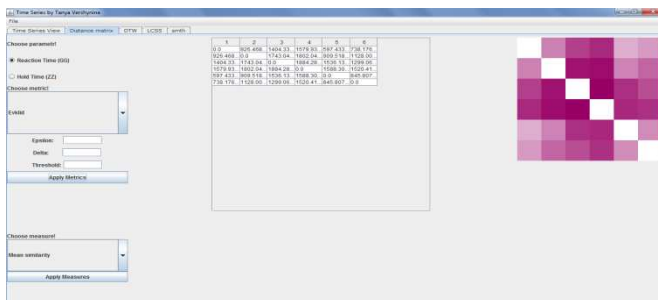


Рисунок 3 – Інтерфейс програмного забезпечення

Висновки. Розроблено інформаційну технологію аналізу динаміки швидкості реакції операторів на основі методів порівняння часових рядів (DTW, LCSS, модифікованого LCSS, TQuest, Евклідової метрики та на основі коефіцієнта кореляції Пірсона, середньої, середньої квадратичної та пікової міри подібності, показників тренду та сезонності), побудови дендрограми методом ієрархічної кластеризації; здійснення сегментації багатовимірних рядів. Створено програмне забезпечення, що реалізує ці алгоритми, забезпечує візуалізацію даних і отриманих результатів для зручного аналізу.

Бібліографічні посилання

1. Esling P., Agon C. Time-series data mining // ACM Computing Surveys (CSUR). 2012. Vol. 45, No 1. Article 12. 34р.
2. Wang X., Mueen A., Ding H., Trajcevski G., Scheuermann P., Keogh E. Experimental comparison of representation methods and distance measures for time series data // Data Mining and Knowledge Discovery. 2013. Vol. 26, No 2. P. 275–309.
3. Iglesias F., Kastner W. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns // Energies. 2013. Vol. 6. P. 579–597.
4. Lovrić M., Milanović M., Stamenković M. Algorithmic methods for segmentation of timeseries: an overview // JCEBI. 2014. Vol.1. No.1 P. 31–53.
5. Keogh E., Chu S., Hart D., Pazzani M. Segmenting Time Series: A Survey and Novel Approach // Data Mining in Time Series Databases. 2004. P. 1-21.
6. Graves D., Pedrycz W. Greedy Gaussian Segmentation of Multivariate Time Series URL: <https://pdfs.semanticscholar.org/6458/cc3830bd7bd3137f30936be69fb2e6cf2cbe.pdf>

7. Graves D., Pedrycz W. Multivariate Segmentation of Time Series with Differential Evolution URL: <https://arxiv.org/abs/1610.07435>

8. Abonyi J., Feil B., Nemeth S., Arva P., Modified Gath-Geva clustering for fuzzy segmentation of multivariate time-series // Fuzzy Sets and Systems. 2005. 149(1) P. 39–56.

9. Жамбю М. Иерархический кластер-анализ и соответствия. М. : Финансы и статистика, 1988. 342 с.

Надійшла до редколегії 05.07.18