

УДК 004.94

DOI: 10.31891/CSIT-2020-2-6

БОРОВИК О. В.

Національна академія Державної прикордонної служби України ім. Богдана Хмельницького, м. Хмельницький

БОРОВИК Д. О., КОСТЕЛЬНА Т. В.

Хмельницький національний університет

ЩОДО НЕОБХІДНОСТІ УДОСКОНАЛЕННЯ МЕТОДУ КЛАСТЕРИЗАЦІЇ МАРШРУТІВ СУДЕН ЯК ПРОЦЕДУРНОГО МОДУЛЯ АВТОМАТИЗОВАНОЇ СИСТЕМИ ОБРОБКИ ДАНИХ У СИСТЕМІ ВИСВІТЛЕННЯ НАДВОДНОЇ ОБСТАНОВКИ

Стаття присвячена аналізу існуючого методу кластеризації маршрутів суден в межах виключної морської (економічної) зони з позиції обґрунтування можливості його застосування або визначення шляхів його удосконалення для використання в системі висвітлення надводної обстановки в інтересах забезпечення національної безпеки на державному кордоні.

У результаті проведеного дослідження встановлено, що існуючий метод кластеризації маршрутів суден не може у явному вигляді бути застосовним для використання в системі висвітлення надводної обстановки як процедурний модуль автоматизованої системи обробки даних, який би забезпечував достатній рівень достовірності виявлення ознак порушення прикордонного законодавства в межах виключної морської (економічної) зони. Також встановлено, що удосконалення методу кластеризації маршрутів суден має стосуватися наступних питань: визначення таких видів апроксимації для формування неперервних маршрутів руху окремих суден, які б забезпечували достатній рівень достовірності результатів на заданих експериментальних наборах даних, що характеризують координати дискретних точок місця знаходження суден в окремі моменти часу; обґрунтування вибору такої кількості експериментальних точок місць знаходження суден і безпосередньо комбінації точок, які б забезпечили достатній рівень достовірності висновків, що формуються; дослідження різних метрик для встановлення подібності довільних маршрутів руху суден з позиції забезпечення однозначності висновків; адаптації запропонованого інструментарію на граничні початкові умови досліджуваної задачі; адаптації методу для випадку довільної складності маршруту руху суден; формування такого методу побудови еталонного маршруту у межах шуканого кластеру, який би забезпечував відтворення маршруту у вигляді неперервної траєкторії.

Ключові слова: система висвітлення надводної обстановки, процедурний модуль, автоматизована система обробки даних, метод, кластеризація, маршрут.

BOROVYK O.

The National Academy of the State Border Guard Service of Ukraine named after Bohdan Khmelnytskyi
Khmelnytskyi city

BOROVYK D., KOSTELNA T.

Khmelnytsky National University

ON THE NECESSITY OF IMPROVING THE METHOD OF CLUSTERIZATION OF VESSEL ROUTES AS A PROCEDURAL MODULE OF THE AUTOMATED DATA PRODUCTION PROCESSING SYSTEM IN SUN

The article is devoted to the analysis of the existing method of clustering of ship routes within the exclusive sea (economic) zone from the standpoint of substantiation of its application or identification of ways to improve it for use in the surface lighting system in the interests of national security at the state border.

The study found that the existing method of clustering ship routes can not be explicitly applicable for use in the surface lighting system as a procedural module of the automated data processing system, which would provide a sufficient level of reliability to detect signs of violations of border legislation within the exclusive maritime (economic) zone. It is also established that the improvement of the method of clustering of routes of ships should address the following issues: definition of such types of approximation for the formation of continuous routes of individual ships, which would provide a sufficient level of reliability of the results on given experimental data sets. time; justification for the choice of such a number of experimental points of location of ships and directly a combination of points that would ensure a sufficient level of reliability of the conclusions formed; study of different metrics to establish the similarity of arbitrary routes of ships from the standpoint of ensuring unambiguous conclusions; adaptation of the proposed tools to the boundary initial conditions of the studied problem; adaptation of the method for the case of arbitrary complexity of the route of vessels; formation of such a method of constructing a reference route within the desired cluster, which would ensure the reproduction of the route in the form of a continuous trajectory.

Keywords: surface lighting system, procedural module, automated data processing system, method, clustering, route.

Постановка проблеми. Важливим чинником стабільності держави є надійна охорона кордону. В сучасних умовах постійного зростання обсягів даних, які необхідно враховувати при прийнятті рішень, надзвичайно важливою є інформаційна складова сучасної моделі охорони кордону. Ця складова реалізується з використанням інтегрованої інформаційно-телекомунікаційної системи (ІТС) «Гарт». Однією з важливих складових ІТС «Гарт» є інформаційно-телекомунікаційна система (ІТС) Морської охорони «Гарт-12».

На сьогоднішній «Гарт-12» використовується лише для висвітлення надводної обстановки в Азово-Чорноморській акваторії. При цьому, існує можливість отримання інформації з ІТС «Гарт-12» про поточне положення суден, порт відправки та порт призначення, тип вантажу та інші додаткові дані. При виборі судна у системі висвітлення надводної обстановки (СВНО) існує можливість перегляду попередніх точок його маршруту, отриманих в певні дискретні моменти часу. При цьому, візуалізація маршруту здійснюється з використанням лінійної апроксимації (наявні точки з'єднуються відрізками прямих) [1].

Слід відмітити, що по такій, наявній у межах СВНО інформації, навіть експертам може бути важко визначити потенційні загрози, що стосуються порушення прикордонного законодавства. Оскільки ж відхилення судна від еталонного маршруту в окремих випадках може являти собою порушення правил прикордонного режиму, то актуальним завданням є ідентифікація тих ознак, які можуть вказувати на зазначене порушення.

Виявлення цих ознак є достатньо складним завданням, оскільки пов'язане з необхідністю обробки значних обсягів даних щодо надводної обстановки в ІТС «Гарт-12» та складністю аналізу ризиків. Саме цим пояснюється той факт, що для вирішення цього завдання авторами роботи [1] запропоновано підхід, що стосується створення відповідної автоматизованої системи обробки даних, яка б містила процедури кластеризації маршрутів, встановлення еталонних маршрутів руху суден в СВНО, виявлення аномалій руху суден і оцінки ризиків. Зрозуміло, що від ефективності функціонування цієї системи безпосередньо залежить ефективність СВНО.

Разом з тим, незважаючи на наявність підходів до комплексної оцінки ефективності функціонування СВНО, загалом [2], на сьогодні ще залишається невирішеним питання оцінки ефективності автоматизованої системи обробки даних. Оскільки остання є інтелектуальною складовою СВНО, то питання оцінки її ефективності є особливо актуальним завданням, яке потребує вирішення.

Аналіз останніх досліджень і публікацій. Останнє передбачає, насамперед, проведення аналізу науково-методичних підходів до розбудови вказаної автоматизованої системи обробки даних.

Відповідні питання знайшли відображення в ряді наукових праць, зокрема [1–6]. Так, у роботі [1] розглянуті можливі підходи до просторового аналізу даних про маршрути суден в ІТС Морської охорони «Гарт-12». Показано доцільність введення особливої метрики для класифікації маршрутів, які представлені в ІТС «Гарт-12» впорядкованими множинами точок, що отримані в довільні моменти часу. Для випадку лінійної апроксимації даних отримано вирази для встановлення ступеню подібності маршрутів. Однак питанням адекватності запропонованої метрики у цій праці увага приділена неналежна.

Логічним продовженням вказаного дослідження стали формалізація та алгоритмізація методики визначення відстаней між маршрутами суден в «Гарт-12». Відповідні питання знайшли відображення у працях [3, 4]. Так, у роботі [3] деталізована методика обчислення метрики для встановлення міри подібності маршрутів суден на основі даних в ІТС морської охорони «Гарт-12». Перевірено комутативність цієї метрики. З метою класифікації маршрутів розроблено метод кластеризації, який дозволяє встановивши належність даних про окремі маршрути до кластерів, визначити еталонний маршрут. Подібний підхід досліджувався і в праці [4]. Проте питанням адекватності методики перевірки комутативності метрики у цих працях увага не приділена взагалі.

Наявність методики кластеризації наявних даних про маршрути суден дозволила авторам роботи [5] запропонувати підхід до визначення просторових аномалій руху суден та оцінки ступеня ризиків порушення прикордонного законодавства. На основі кластеризації було запропоновано технологію проведення класифікації наявних даних у базі ІТС «Гарт-12» та визначення еталонних маршрутів. Дані про ці маршрути дозволяють класифікувати нові дані, які надходять до СВНО, та проводити на їх основі аналіз ризиків. Також у роботі визначено логічні вирази для встановлення рівня загроз щодо порушення прикордонного законодавства на морській ділянці кордону.

Таким чином, з проведеного аналізу випливає, що базовим положенням, яке визначає ефективність автоматизованої системи обробки даних, є метод кластеризації маршруту суден. Саме цей метод є інтелектуальною основою можливості виявлення ознак порушення прикордонного законодавства. А отже, саме достовірність результату, отриманого за допомогою цього методу, здійснює визначальний вплив на ефективність досліджуваної автоматизованої системи.

Цим і визначається необхідність проведення оцінки достовірності результатів, які формуються методом кластеризації маршрутів суден.

Метою статті є аналіз методу кластеризації маршрутів суден з позиції обґрунтування можливості його застосування або визначення шляхів його удосконалення для використання в СВНО як процедурного модуля автоматизованої системи обробки даних, який би забезпечував достатній рівень достовірності виявлення ознак порушення прикордонного законодавства.

Виклад основного матеріалу дослідження. У базі даних ІТС морської охорони «Гарт-12» інформація про пересування суден зберігається у вигляді впорядкованої послідовності точок з координатами суден у певні моменти часу.

Уявлення про місце знаходження суден у фіксований момент часу можна сформулювати, наприклад, з використанням рис. 1.

Сукупність координат точок місцезнаходження i -го судна можна позначити впорядкованою множиною (кортежем)

$$M_i = \{(x_1, y_1), \dots, (x_{k_i}, y_{k_i})\}, \quad (1)$$

де k_i – потужність множини M_i .

Слід зауважити, що потужності відповідних множин для різних суден, у загальному випадку, різні. Причому потужності цих множин можуть відрізнятися достатньо суттєво навіть у випадку, якщо судна

рухаються по одному маршруту між заданими пунктами відправлення та призначення. Траєкторії руху суден у вигляді неперервних кривих відсутні.

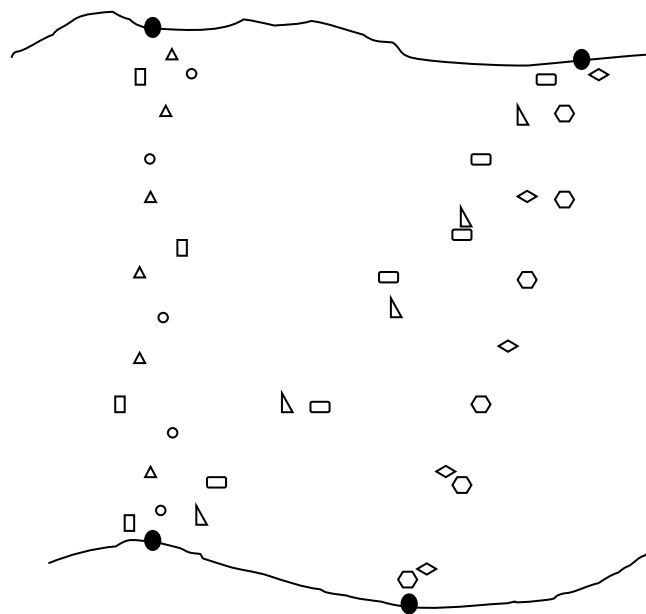


Рис. 1. Схематичне представлення даних щодо положення окремих суден у СВНО:
 чорні точки — пункти відправлення суден;
 $\square \triangle \diamond \circ \square$ — позначення положення окремих суден

Дослідження питання про можливість порушення судном правил прикордонного режиму передбачає вирішення комплексу взаємопов'язаних задач:

1. Встановлення подібності маршрутів довільних пар різних суден;
2. Встановлення кластеру маршрутів між заданими пунктами відправлення та призначення (транспортного коридору між двома пунктами);
3. Визначення еталонного маршруту між двома пунктами в межах встановленого кластеру;
4. Встановлення належності траєкторії руху судна кластеру маршрутів між заданими пунктами відправлення та призначення;
5. Оцінка ступеню відхилення траєкторії досліджуваного судна від еталонного маршруту.

У межах даної роботи аналізу піддаватимуться лише перші дві задачі.

Задача 1. Нехай задано маршрути руху деякої пари суден у вигляді множин M_1, M_2 типу (1).

Ступінь подібності маршрутів M_1 та M_2 (див. рис. 2) у роботі [1] запропоновано визначати за допомогою метрики $R(M_1, M_2)$.

Для розрахунку $R(M_1, M_2)$ необхідно насамперед здійснити встановлення (відновлення, генерування) неперервного маршруту руху першого судна між заданими у дискретній множині M_1 точками з використанням лінійної апроксимації (див. рис. 2). Величина $R(M_1, M_2)$ визначається, як середнє відхилення точок множини M_2 , які описують дискретний маршрут руху другого судна, від апроксимованого неперервного першого маршруту у вигляді:

$$R(M_1, M_2) = \frac{\sum_{j=1}^{k_2} R_{\min j}}{k_2}, \quad (2)$$

де $R_{\min j}$ є найкоротшою відстанню від j -ї точки множини M_2 до встановленого першого маршруту, тобто до найближчого деякого i -го сегменту прямої, який з'єднує точки (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_1 , а k_2 — кількість досліджуваних точок другого маршруту.

При використанні лінійної апроксимації сегментів першого маршруту координати (x, y) точки i -го сегменту прямої, який з'єднує точки (x_i, y_i) і (x_{i+1}, y_{i+1}) першого маршруту, яка є найближчою до j -ї точки множини M_2 , координати якої рівні (x_j, y_j) , можна отримати з формул:

$$x = \frac{y_j \cdot k_i + x_j - b_i \cdot k_i}{1 + k_i^2}, \quad (3)$$

$$y = k_i \cdot x + b_i,$$

де k_i, b_i – параметри лінійного рівняння прямої, що описує i -й сегмент, який з'єднує точки (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_1 . Вказані параметри можуть бути нескладно знайдені з використанням рівняння прямої, що проходить через дві задані точки.

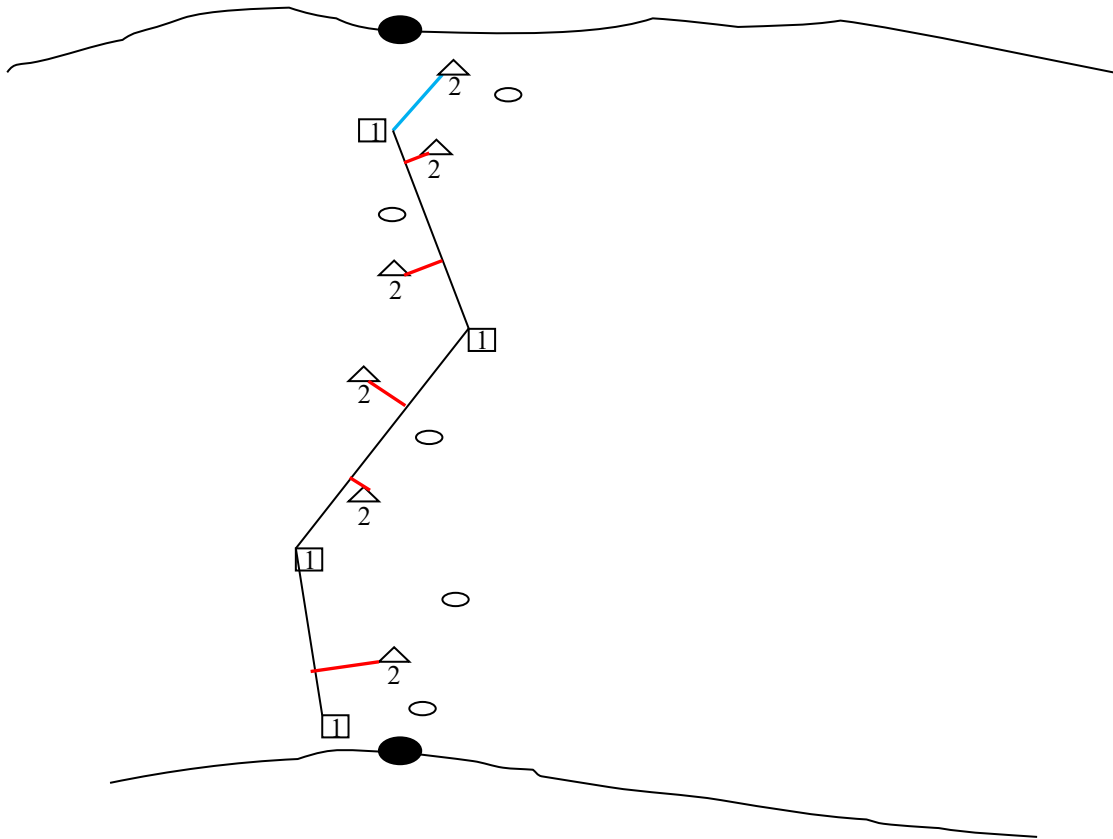


Рис.2. Підхід до визначення ступеня подібності маршрутів M_1 та M_2 за допомогою метрики $R(M_1, M_2)$

Тоді

$$R_{\min j} = \sqrt{\left(x_j - \frac{y_j \cdot k_i + x_j - b_i \cdot k_i}{1 + k_i^2}\right)^2 + \left(y_j - k_i \cdot \frac{y_j \cdot k_i + x_j - b_i \cdot k_i}{1 + k_i^2} - b_i\right)^2}. \quad (4)$$

Формула (4) може застосовуватись у випадку, коли найкоротша відстань від j -ї точки множини M_2 до встановленого першого маршруту знаходиться на перпендикулярі від j -ї точки множини M_2 до деякого сегмента неперервного апроксимованого першого маршруту (на рис. 2 це відрізки, що зображені червоним кольором). Однак, можливий випадок, коли величина x , що отримана з (3), може знаходитись поза діапазоном $[x_i, x_{i+1}]$. Тоді величину $R_{\min j}$ пропонується знаходити як найменшу відстань від j -ї точки множини M_2 до кінців відрізка i -го сегменту (тобто до точок (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_1) (на рис. 2 це відрізок, що зображений синім кольором). Тоді, з урахуванням цього, у загальному випадку:

$$R_{\min j} = \begin{cases} \sqrt{\left(x_j - \frac{y_j \cdot k_i + x_j - b_i \cdot k_i}{1 + k_i^2}\right)^2 + \left(y_j - k_i \cdot \frac{y_j \cdot k_i + x_j - b_i \cdot k_i}{1 + k_i^2} - b_i\right)^2}, & x \in [x_i, x_{i+1}], \\ \min \left\{ \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}, \sqrt{(x_j - x_{i+1})^2 + (y_j - y_{i+1})^2} \right\}, & x \notin [x_i, x_{i+1}]. \end{cases} \quad (5)$$

Послідовне використання виразів (5), (3), (2) визначає методику обчислення метрики $R(M_1, M_2)$.

Слід зауважити, що величина метрики R , яка використовується для оцінки ступеня подібності першого та другого маршрутів залежить від порядку слідування аргументів у записі для R . Так, якщо в записі $R(M_1, M_2)$ аргументи M_1 та M_2 поміняти місцями, то запис метрики матиме вигляд $R(M_2, M_1)$, а її відшукування для того ж випадку початкових даних, що й наведені на рис. 2, може бути оцінене з рис. 3.

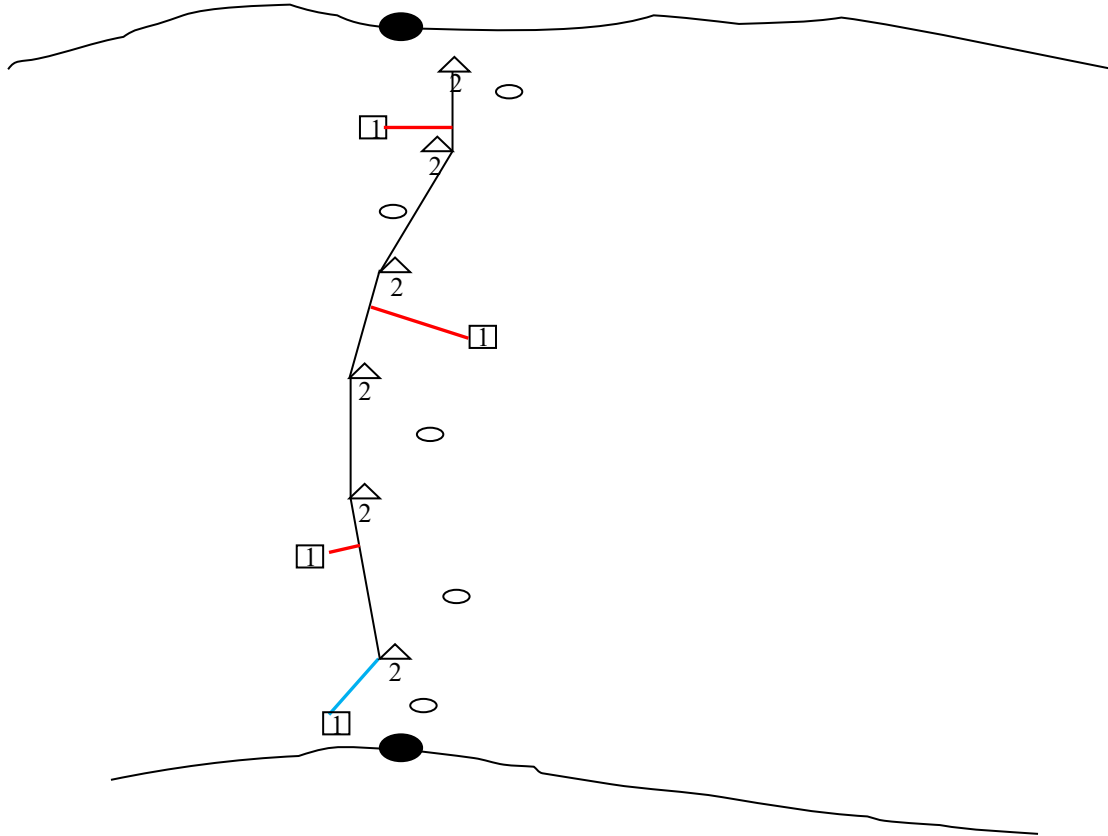


Рис. 3. Підхід до визначення ступеня подібності маршрутів M_2 та M_1 за допомогою метрики $R(M_2, M_1)$

Для розрахунку $R(M_2, M_1)$ необхідно насамперед здійснити встановлення (відновлення, генерування) неперервного маршруту руху другого судна між заданими у дискретній множині M_2 точками з використанням лінійної апроксимації (див. рис 3). Тоді величина $R(M_2, M_1)$ визначається, як середнє відхилення точок множини M_1 , які описують дискретний маршрут руху першого судна, від апроксимованого неперервного другого маршруту у вигляді:

$$R(M_2, M_1) = \frac{\sum_{j=1}^{k_1} R_{\min j}}{k_1}, \quad (6)$$

де $R_{\min j}$ є найкоротшою відстанню від j -ї точки множини M_1 до встановленого другого маршруту, тобто до найближчого деякого i -го сегменту прямої, який з'єднує точки (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_2 , а k_1 – кількість досліджуваних точок першого маршруту.

При використанні лінійної апроксимації сегментів другого маршруту координати (x, y) точки i -го сегменту прямої, який з'єднує точки (x_i, y_i) і (x_{i+1}, y_{i+1}) другого маршруту, яка є найближчою до j -ї точки множини M_1 , координати якої рівні (x_j, y_j) , також можна отримати за допомогою формул (3), але з урахуванням того, що k_i, b_i – параметри лінійного рівняння прямої, що описує i -й сегмент, який з'єднує

точки (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_2 . Вказані параметри можуть бути нескладно знайдені з використанням рівняння прямої, що проходить через дві задані точки. Тоді $R_{\min j}$ можна визначати за допомогою формули (4) у випадку, коли найкоротша відстань від j -ї точки множини M_1 до встановленого другого маршруту знаходиться на перпендикулярі від j -ї точки множини M_1 до деякого сегмента неперервного апроксимованого другого маршруту (на рис. 3 це відрізки, що зображені червоним кольором).

У випадку ж, коли величина x , що отримана з (3), знаходиться поза діапазоном $[x_i, x_{i+1}]$, величину $R_{\min j}$ пропонується знаходити як найменшу відстань від j -ї точки множини M_1 до кінців відрізка i -го сегменту (тобто до точок (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_2) (на рис. 3 це відрізок, що зображений синім кольором). Тоді, з урахуванням цього, у загальному випадку $R_{\min j}$ буде також знаходитись за допомогою формули (5).

Аналіз підходів і даних, що наведені на рис. 2, 3, спонукає до необхідності перевірки рівності метрик $R(M_1, M_2)$, $R(M_2, M_1)$. Адже, для однієї і тієї ж пари маршрутів у залежності від обраної метрики у разі неспівпадання $R(M_1, M_2)$ і $R(M_2, M_1)$ можуть мати місце різні висновки щодо подібності чи неподібності маршрутів. Саме тому для перевірки рівності величин $R(M_1, M_2)$ і $R(M_2, M_1)$ у роботі [1] було проведено окреме дослідження для вихідних даних, які наведені на рис. 4.

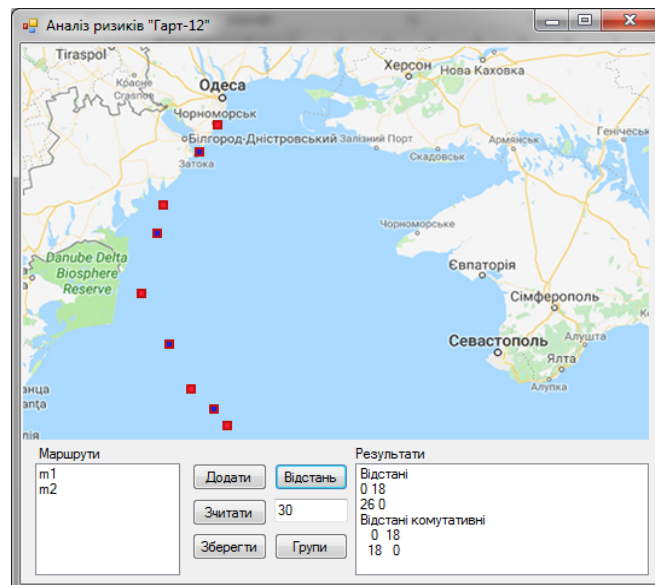


Рис. 4. Розрахунок метрик $R(M_1, M_2)$ і $R(M_2, M_1)$

Застосування формул (2) – (6) для початкових даних рис. 4 дозволило отримати, що $R(M_1, M_2) = 18$, $R(M_2, M_1) = 26$. А отже, у загальному випадку, дійсно $R(M_1, M_2) \neq R(M_2, M_1)$.

Відсутність властивості комутативності щодо метрик $R(M_1, M_2)$ і $R(M_2, M_1)$ спонукала авторів роботи [1] до формування пропозицій щодо застосування нової метрики, яка б не містила недоліку, що стосувався б неоднозначності формування висновків відносно подібності маршрутів.

Ця пропозиція має вигляд метрики виду:

$$R^k(M_1, M_2) = \min(R(M_1, M_2), R(M_2, M_1)). \quad (7)$$

Для перевірки властивості комутативності для метрики (7) було проведено дослідження для початкових даних, що наведені на рис. 5. Результати дослідження, що описані у праці [3], вказують на комутативність метрики (7).

Однак питанню об'єктивності результатів, отриманих за допомогою метрики (7), у роботі [3] увага не приділялася.

Задача 2. Введення деякого граничного порогового рівня r_{pk} (даний рівень визначається на основі експертної оцінки) щодо величини відхилення маршрутів і його порівняння з метрикою (7) дозволяє робити висновки про ступінь подібності маршрутів M_1 і M_2 незалежно від того, для якого з маршрутів формується апроксимаційна неперервна траєкторія руху, і відносно якого здійснюється оцінка близькості до такої траєкторії.

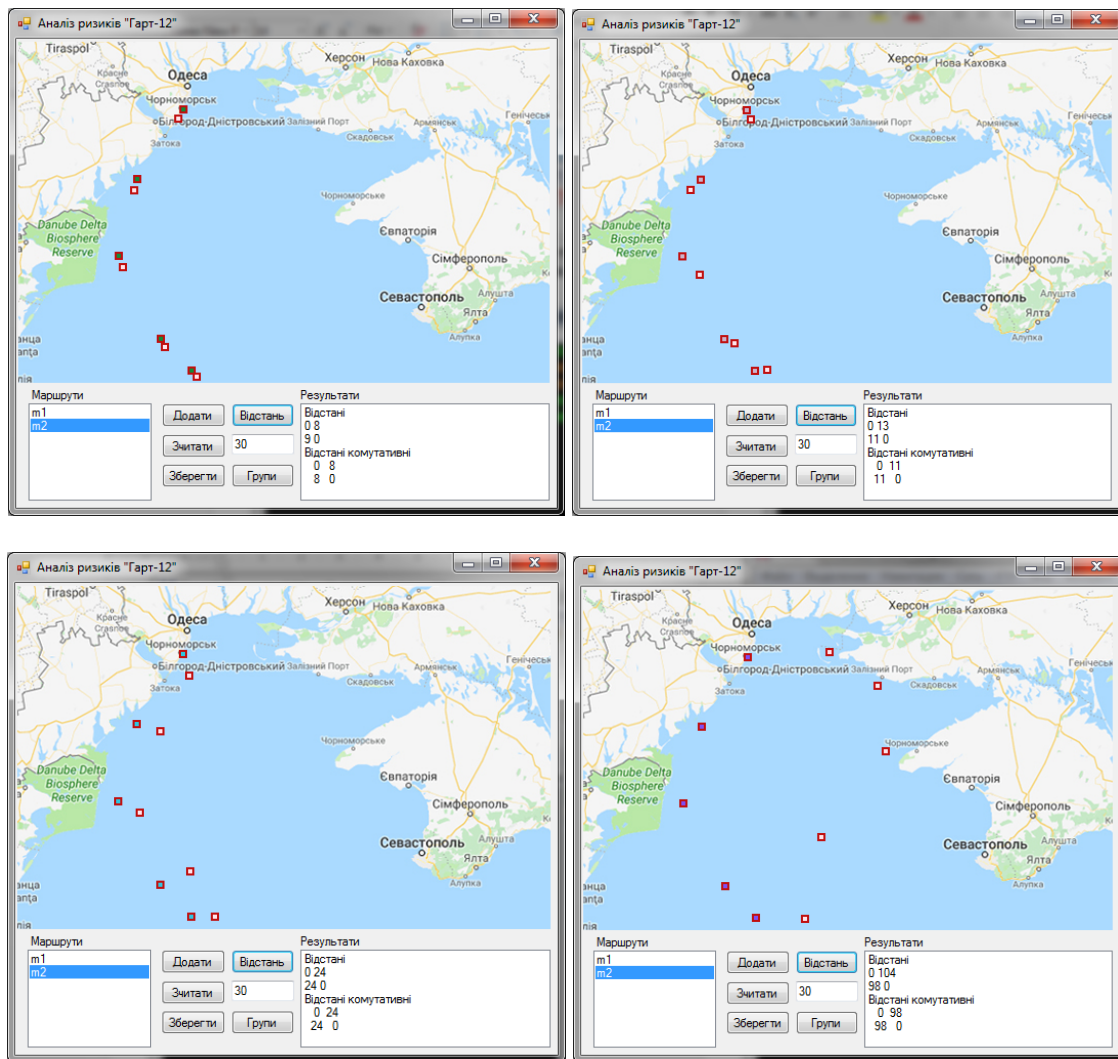


Рис. 5. Приклади обчислення метрики (7)

Встановлення геометричної близькості маршрутів за допомогою формули (7) дає можливість провести їх подальшу класифікацію та поділ на кластери (кластеризацію) з подальшим визначенням у кожному з них еталонного маршруту. Еталонні маршрути можуть використовуватись для оцінки ступеня відхилення судна, відносно якого проводиться аналіз. Також може здійснюватись подальша класифікація нових даних, що поступають до системи, та оновлення або уточнення меж кластера.

Для проведення кластеризації маршрутів на основі наявних у базі «Гарт-12» даних у роботі [3] пропонується побудувати матрицю \mathbf{R} , елементи r_{ij} якої визначаються наступним чином:

$$r_{ij} = \begin{cases} 0, & i = j, \\ R^k(M_i, M_j), & i \neq j, \end{cases} \quad (8)$$

де M_i, M_j – маршрути відповідно i - та j -го суден.

Після формування матриці \mathbf{R} для кожного її рядка, тобто для $i = \text{const}$, обчислюється:

– кількість n_s елементів, які не перевищують порогового для кластеризації рівня r_{pk} , тобто кількість елементів, для яких у рядку i виконується нерівність $R^k(M_i, M_j) \leq r_{pk}$;

– величина $s_v = \sum R^k(M_i, M_j)$, яка формується лише з тих значень $R^k(M_i, M_j)$, для яких виконується умова $R^k(M_i, M_j) \leq r_{pk}$.

На основі отриманих даних, які сортуються за зменшенням n_s (для елементів з однаковими значеннями n_s здійснюється сортування за збільшенням s_v), формується впорядкована множина \mathbf{M}_s , елементами якої є набори $(i, n_s(i), s_v(i))$.

Далі множина M_s використовується для визначення тих маршрутів, які можуть бути включені до одного і того ж кластера. Здійснюється це за допомогою наступного алгоритму:

1. Якщо для першого елемента множини M_s виконується умова $n_s(i) > 0$, то створюється кластер з центральним маршрутом i і в нього включаються всі маршрути j , для яких $R^k(M_i, M_j) \leq r_{pk}$.

2. З множини M_s видаляються всі набори $(i, n_s(i), s_v(i))$, в яких перший елемент i співпадає із значеннями j , які визначили маршрути, що увійшли до створеного кластера.

3. Кроки 1–2 алгоритму повторюються до тих пір, поки потужність множини M_s є більшою за 0.

Реалізація такого підходу дозволяє встановити всі кластери (коридори руху суден), які задовольняють умові не перевищення порогу r_{pk} .

Проведений вище аналіз задач 1, 2 дозволяє зробити висновки про те, що у них є багато моментів, які можуть впливати на достовірність результатів моделювання та зменшувати ефективність автоматизованої системи обробки даних СВНО.

До числа таких можна віднести:

1. Припущення про доцільність застосування лінійної апроксимації при роботі з окремими сегментами між точками множини $M_i = \{(x_1, y_1), \dots, (x_{ki}, y_{ki})\}$ при формуванні неперервного апроксимаційного маршруту чи M_1 , чи M_2 , є сумнівним і нічим не підтвердженим. Особливий сумнів виникає у випадку значної віддаленості в часі або просторі досліджуваних точок. А тому актуальності набуває завдання дослідження достовірності результатів на заданих експериментальних наборах даних у разі, якщо застосовуватиметься інший вид апроксимації (в тому числі за допомогою сплайнів). При цьому, окремої уваги потребує питання відповідності обраного методу апроксимації логічному змісту процесу руху суден.

2. Дослідження маршрутів різних суден на предмет формування кластерів може здійснюватись не лише в межах між пунктами відправлення та призначення, а й між довільними двома точками, в межах яких існує особлива необхідність вивчення питання про порушення прикордонного законодавства. У цьому випадку застосування формули (2) для відшукування метрики $R(M_1, M_2)$, формули (6) для відшукування метрики $R(M_2, M_1)$, або формули (7) для встановлення метрики $R^k(M_1, M_2)$, потребує обґрунтування вибору кількостей k_2 і k_1 необхідних для дослідження точок, які б забезпечили достатній рівень достовірності висновків, що формуються.

3. Для умов, що наведені вище, актуальним є питання і вибору для дослідження конкретних точок на маршрутах M_1 та M_2 , які при реалізації значень k_2 і k_1 забезпечували б достатній рівень достовірності висновків, що формуються. Актуальність дослідження цього питання обумовлюється тим, що при реалізації різних комбінацій досліджуваних точок величина $R(M_1, M_2)$, або $R(M_2, M_1)$, або $R^k(M_1, M_2)$ може суттєво відрізнятись, що може впливати на достовірність методики кластеризації.

4. Викликає сумнів і питання забезпечення об'єктивності результатів, що визначаються застосуванням метрики (7). Адже властивість комутативності метрики справедлива не лише у випадку застосування (7). Так, при застосуванні метрики:

$$R^k(M_1, M_2) = \max(R(M_1, M_2), R(M_2, M_1)), \quad (9)$$

або

$$R^k(M_1, M_2) = \frac{R(M_1, M_2) + R(M_2, M_1)}{2}, \quad (10)$$

або

$$R^k(M_1, M_2) = \sqrt{R(M_1, M_2)R(M_2, M_1)}, \quad (11)$$

існує велика ймовірність її комутативності, а отже, можливості застосування. Однак висновки щодо подібності маршрутів та співпадання меж кластерів при застосуванні наведених метрик у різних випадках можуть відрізнятись. А отже, питання адекватності застосовуваного інструментарію та достовірності результатів, що ним забезпечуються, потребують окремого вивчення.

5. При застосуванні метрик (2), або (6), або (7) може мати місце випадок, коли для окремих сегментів між точками (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_1 або M_2 виконуватиметься умова $x_i = x_{i+1}$. Тобто у вибраній системі координат окремі сегменти можуть бути паралельними осі ординат. У цьому випадку формула (5) не може бути застосовною для відшукування величини $R_{\min j}$.

6. У разі, якщо деякий досліджуваний маршрут M_i на окремій ділянці маршруту між пунктом відправлення та призначення паралельний еталонному маршруту M_{et} , але на цій ділянці величина

$R(M_i, M_{et}) > r_{pk}$, то це може спричинити не включення маршруту M_i у кластер, хоча у загальному випадку цей маршрут може увійти до цього кластеру. Разом з тим, існуюча методика кластеризації цієї особливості не враховує. А оскільки така особливість важлива з позиції формування висновків щодо наявності порушень прикордонного законодавства, то відповідне питання також потребує окремого дослідження.

7. У разі достатньої складності маршруту існуюча методика може бути незастосовною, оскільки може не відповідати логічному змісту досліджуваного процесу.

8. Формування еталонного маршруту M_{et} описане у роботі [5]. Цей маршрут являє собою набір дискретних точок. Актуальним є питання формування такого методу побудови еталонного маршруту M_{et} , який би забезпечував створення еталонного маршруту у вигляді неперервної траєкторії.

Висновки. Проведене дослідження дозволяє зробити висновок про те, що базовим положенням, яке дійсно визначає ефективність автоматизованої системи обробки даних, є метод кластеризації маршруту суден. Існуючий метод кластеризації характеризується рядом особливостей, які наведені вище та стосуються інструментальних особливостей реалізації і впливу на достовірність результатів для подальшої оцінки ризиків протиправної діяльності на державному кордоні. Аналіз визначених особливостей дозволяє зробити висновок про те, що існуючий метод кластеризації маршрутів суден не може у явному вигляді бути застосовним для використання в СВНО як процедурний модуль автоматизованої системи обробки даних, який би забезпечував достатній рівень достовірності виявлення ознак порушення прикордонного законодавства.

Напрямами подальших досліджень вбачається вивчення питань, які окреслені проблемними моментами 1-8 застосування існуючого методу кластеризації, а також їх подальша формалізація, алгоритмізація, програмування та апробація.

Література

1. Боровик О. В., Рачок Р. В., Мазур В. Ю. Визначення підходів до просторового аналізу даних в інформаційно-телекомунікаційній системі морської охорони «Гарт-12» // Збірник наукових праць № 1 (75). Серія: Військові та технічні науки. – Хмельницький: Вид. НАДПСУ, 2018. – С. 134-143.
2. Мазур В. Ю., Боровик О. В. Методичні основи оцінки ефективності функціонування єдиної системи висвітлення надводної обстановки на морській ділянці. – Наука і техніка Повітряних Сил Збройних Сил України № 2 (31). – Харків: ХНУПС, 2018. – С. 182-189.
3. Мазур В. Ю., Боровик О. В., Рачок Р. В. Метод кластеризації маршрутів суден в системі висвітлення надводної обстановки // Сучасні інформаційні технології у сфері безпеки та оборони Вип. № 2(32). – К.: Національний університет оборони України імені Івана Черняхівського, Інститут інформаційних технологій, 2018. – С. 87-92.
4. Pan Sheng, Jingbo Yin Extracting Shipping Route Patterns by Trajectory Clustering Model Based on Automatic Identification System Data. - Sustainability 2018, 10.
5. Боровик О. В., Рачок Р. В., Мазур В. Ю. Методика виявлення просторових аномалій руху суден та її використання при оцінюванні ризиків у системі морської охорони «Гарт-12» // Наука і оборона. – 2018. - № 2. С. 65-69.
6. Arguedas V. F., Pallotta G., Vespe M. Maritime Traffic Networks: From Historical Positioning Data to Unsupervised Maritime Traffic Monitoring. - Transactions on intelligent transportation systems, Vol. 19, № 3, March 2018. - P. 722-732.

References

1. Borovyk O.V., Rachok R.V., Mazur V.Yu. Definition of approaches to spatial data analysis in the information and telecommunication system of maritime protection "Gart-12" // Collection of scientific works № 1 (75). Series: Military and technical sciences. - Khmelnytsky: Ed. NADPSU, 2018. - P. 134-143.
2. Mazur V. Yu., Borovyk O.V. Methodical bases of an estimation of efficiency of functioning of uniform system of illumination of a surface situation on a sea site. - Science and Technology of the Air Force of the Armed Forces of Ukraine № 2 (31). - Kharkiv: KhNUPS, 2018. - P. 182-189.
3. Mazur V. Yu., Borovyk O.V., Rachok R.V. The method of clustering of ship routes in the system of surface lighting // Modern information technologies in the field of security and defense Vol. № 2 (32). - Kyiv: Ivan Chernyakhovsky National University of Defense of Ukraine, Institute of Information Technologies, 2018. - P. 87-92.
4. Pan Sheng, Jingbo Yin Extracting Shipping Route Patterns by Trajectory Clustering Model Based on Automatic Identification System Data. - Sustainability 2018, 10.
5. Borovyk O.V., Rachok R.V., Mazur V.Yu. Methods for detecting spatial anomalies in the movement of ships and its use in risk assessment in the marine security system "Gart-12" // Science and Defense. - 2018. - № 2. - S. 65-69.
6. Arguedas V. F., Pallotta G., Vespe M. Maritime Traffic Networks: From Historical Positioning Data to Unsupervised Maritime Traffic Monitoring. - Transactions on intelligent transportation systems, Vol. 19, № 3, March 2018. - P. 722-732.

Надійшла / Paper received: 22.08.2020
Надрукована / Paper Printed : 03.11.2020