

Валентин Юрійович Мазур (канд. військ. наук, доцент)

Олег Васильович Боровик (доктор техн. наук, професор)

Роман Васильович Рачок (канд. техн. наук, професор)

Національна академія Державної прикордонної служби України імені Богдана Хмельницького, Хмельницький, Україна

МЕТОД КЛАСТЕРИЗАЦІЇ МАРШРУТІВ СУДЕН В СИСТЕМІ ВИСВІТЛЕННЯ НАДВОДНОЇ ОБСТАНОВКИ

У статті вирішена задача класифікації даних про маршрути руху суден в системі висвітлення надводної обстановки. З метою такої класифікації запропоновано використання кластеризації, метод проведення якої розроблений в межах даного дослідження. Особливістю методу є урахування специфіки представлення даних про маршрути руху суден в інформаційно-телекомунікаційній системі морської охорони «Гарт-12». Ця специфіка врахована у метриці та особливостях формування маршрутних кластерів та визначення їх центрів. Такий аналіз даних в системі висвітлення надводної обстановки дозволить вирішувати задачу оцінки ризиків шляхом виявлення просторових аномалій в маршрутах руху суден.

Ключові слова: інформаційно-телекомунікаційна система морської охорони «Гарт-12», система висвітлення надводної обстановки, кластеризація, метрика, маршрут.

Вступ

Специфічним і важливим видом національної безпеки є прикордонна безпека, яка повинна дозволити захищати інтереси особистості, суспільства і держави у прикордонній сфері. Основні завдання в сфері прикордонної безпеки - забезпечення недоторканності державного кордону та охорони суверенних прав України в її виключній (морській) економічній зоні, покладаються на Державну прикордонну службу України (ДПСУ).

Постановка проблеми у загальному вигляді. Ефективна охорона морської ділянки кордону є важливою складовою забезпечення національної безпеки нашої держави. В сучасних умовах ця ефективність значною мірою залежить від своєчасного опрацювання значних обсягів даних про надводну обстановку. Постійне зростання обсягів цих даних, які необхідно врахувати при прийнятті рішень, обумовило останнім часом активний розвиток відповідної інформаційної складової сучасної моделі охорони кордону. Ця складовою реалізована в інформаційно-телекомунікаційній системі (ІТС) морської охорони «Гарт-12» важливим елементом якої є система висвітлення надводної обстановки. Однак, на сьогодні ця система і «Гарт-12» в цілому використовується лише для обліку і відображення даних про рух суден в Азово-Чорноморській акваторії. У межах даної системи існує можливість отримання інформації про поточне положення суден, порт відправки та порт призначення, тип вантажу. При виборі судна у системі висвітлення надводної обстановки існує можливість перегляду попередніх точок його маршруту отриманих в певні дискретні моменти часу. Однак, на основі цих даних визначення потенційних загроз

людиною-експертом та подальше прийняття ефективного рішення є достатньо складним завданням. Допомогти в його вирішенні може створення інтегрованої до ІТС «Гарт-12» системи оцінки ризиків. Одним з аспектів такої оцінки є виявлення просторових аномалій руху, які проявляються у відхиленнях руху суден від типових маршрутів. Однак дискретність даних про маршрути та неспівпадіння у часі окремих відліків суттєво ускладнює цю задачу.

Аналіз останніх досліджень і публікацій, в яких започатковано вирішення даної проблеми та на які опирається автор. Задача виявлення просторових та часових аномалій руху суден була поставлена у дослідженні [1]. Однак при цьому було лише визначено доцільність проведення такого аналізу з використанням геообробки і не визначено можливих напрямів вирішення цього завдання.

Значний розвиток у вирішенні задачі виявлення просторових аномалій був зроблений у дослідженні [2]. У цій роботі було визначено підходи до просторового аналізу даних в інформаційно-телекомунікаційній системі морської охорони «Гарт-12», розроблено метрику порівняння даних про окремі маршрути. Однак запропонована у [2] метрика не позбавлена ряду недоліків. Окрім цього у даному дослідженні лише визначена потреба у класифікації даних про маршрути з використанням введеної метрики, але не сформовано відповідного методу.

Використання класичних методів кластеризації [3-5] для класифікації маршрутів руху морських суден обмежується складністю визначення центрів відповідних кластерів, які представляють собою впорядковані у часі множини точок та

встановлення належності до кластерів окремих маршрутів.

Відповідно до цього, вирішення задачі виявлення просторових аномалій руху суден в системі висвітлення надводної обстановки потребує уточнення відповідної метрики та розробки методу кластеризації маршрутів.

Мета статті – розробка методу кластеризації маршрутів руху суден в системі висвітлення надводної обстановки на основі уточненої метрики.

Виклад основного матеріалу дослідження.

В базі даних інформаційно-телекомунікаційної системи морської охорони «Гарт-12» інформація про пересування суден зберігається у вигляді впорядкованої послідовності точок з координатами суден у певні моменти часу. Сукупність цих координат для *i*-того судна можна позначити впорядкованою множиною (кортежем)

$M_i = \{(x_1, y_1), \dots, (x_{k_i}, y_{k_i})\}$, де k_i – потужність множини M_i [2]. Ступінь подібності маршрутів, які описуються двома такими множинами M_1 та M_2 , можливо визначити на основі введення відповідної метрики $R(M_1, M_2)$.

У дослідженні [2], для розрахунку $R(M_1, M_2)$ запропоноване проводити відновлення першого маршруту між заданими у множині M_1 точками з використанням апроксимації. Тоді величина $R(M_1, M_2)$ визначається, як середнє відхилення

Тоді

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

Однак, можливий випадок, коли величина x , отримана з (3) може знаходитись поза діапазоном $[x_i, x_{i+1}]$. Тоді пропонується знаходити $R_{\min ij}$ як найменшу відстань від j -тої

точок множини M_2 , які описують другий маршрут від відновленого першого маршруту

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

де $R_{\min j}$ є найкоротшою відстанню від j -тої точки множини M_2 до відновленого першого маршруту.

Обчислення $R_{\min j}$ можливо проводити наступним чином

$$R_{\min j} = \min_i (R_{\min ij}), \quad (2)$$

де $R_{\min j}$ – найменша відстань від j -тої точки множини M_2 до i -того сегменту прямої, який поєднує точки (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_1 .

При використанні лінійної апроксимації сегментів першого маршруту, координати (x, y) найближчої точки на прямій до визначеної точки (x_j, y_j) можливо отримати з виразів [2]

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

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

де k_i, b_i – параметри, що описують i -тий сегмент прямої, який поєднує точки (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_1 .

точки множини M_2 до кінців відрізка (точок (x_i, y_i) і (x_{i+1}, y_{i+1}) множини M_1).

Остаточнo

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

Послідовне використання виразів (5), (2), (1) визначає методуку обчислення метрики $R(M_1, M_2)$.

Однак, визначена таким чином метрика не є

комутативною. Тобто $R(M_1, M_2)$ може відрізнитись від $R(M_2, M_1)$. На рисунку 1 представлені результати обчислення $R(M_1, M_2)$ і $R(M_2, M_1)$.

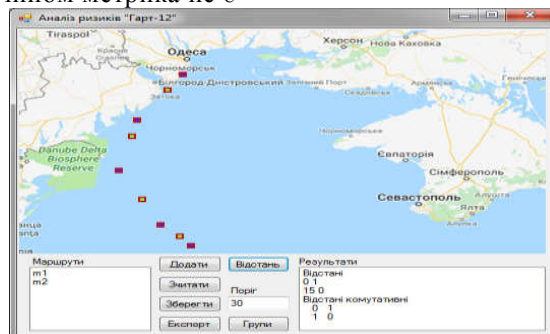


Рис. 1 – Розрахунок метрик $R(M_1, M_2)$ і $R(M_2, M_1)$

Для випадку вихідних даних маршрутів наведених на рисунку 1, $R(M_1, M_2)=1$, $R(M_2, M_1)=15$. Відсутність властивості комутативності пояснюється тим, що при відновленні першого маршруту, точки другого

є близькими до відповідних відрізків. Однак, коли апроксимуються сегменти другого маршруту, точки першого знаходяться на більшій відстані (рисунок 2).

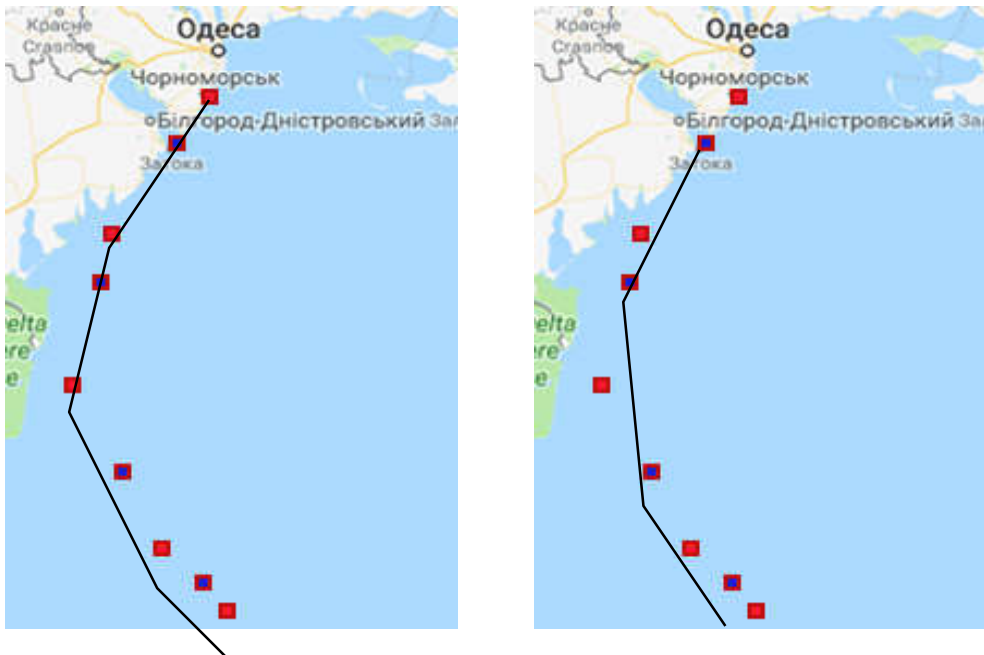


Рис. 2 – Відмінності при апроксимації різних маршрутів

З метою уникнення неоднозначності при визначенні метрики та забезпечення її

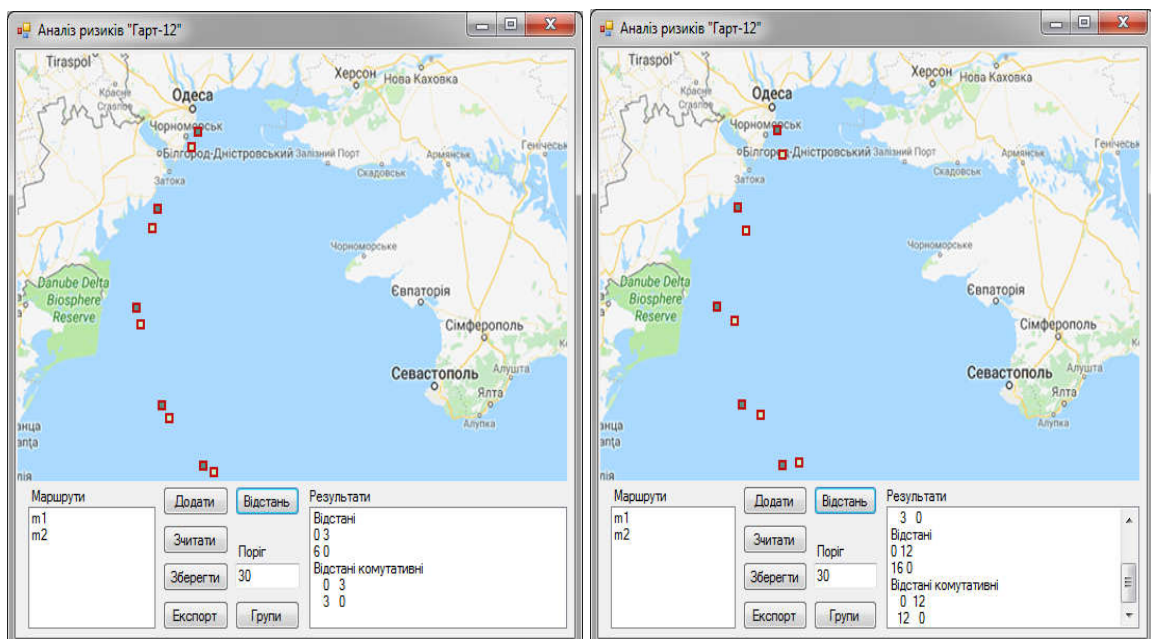
комутативності пропонується відстань між маршрутами обчислювати за виразом

$$R^k(M_1, M_2) = \min(R(M_1, M_2), R(M_2, M_1)) \quad (6)$$

Розглянемо приклади обчислення метрики (6) для різних варіантів маршрутів з поступовим зменшенням їх геометричної подібності (рисунок 3).

отриманих значень (6) показала, що суттєва для класифікації відмінність між маршрутами характерна для випадків, коли вони перевищують 30 км. Однак, ризики порушення прикордонного законодавства судами суттєво зростають при перевищенні значень $R^k(M_1, M_2)$ порогового рівня 10 км.

Як впливає з рисунку 3, при віддаленні дискретних точок представлення двох досліджуваних маршрутів метрика (6) поступово зростає. Експертна оцінка



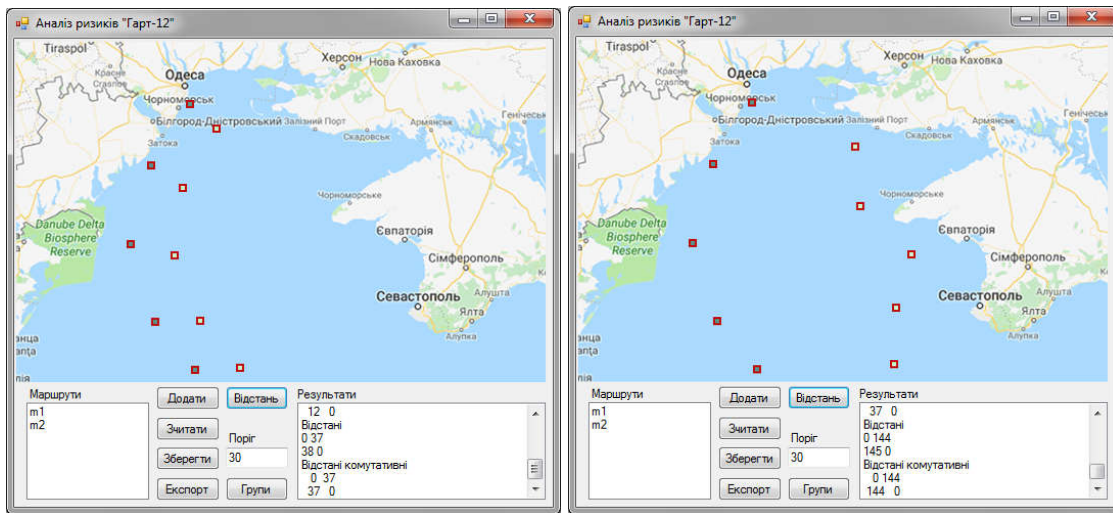


Рис. 3 – Приклади обчислення метрики (6)

Встановлення геометричної близькості маршрутів за (6) дає можливість провести їх подальшу класифікацію та поділ на кластери (кластеризацію) з визначенням центральних в цих кластерах маршрутів. Ці маршрути в подальшому пропонується використовувати як еталонні, відстані від яких можна застосовувати для класифікації нових даних, що поступають до системи і подальшої оцінки ризиків на основі визначення геометричних просторових аномалій руху суден.

Для проведення кластеризації маршрутів на основі наявних у базі «Гарт-12» даних пропонується побудувати матрицю відстаней R . Елементи цієї матриці r_{ij} визначимо наступним чином

$$r_{ij} = \begin{cases} 0, & i = j \\ R^k(M_i, M_j), & i \neq j \end{cases} \quad (7)$$

Після формування матриці R для кожного її рядка i обчислюється:

- кількість елементів у рядку n_s , які не перевищують порогового для кластеризації рівня r_{pk} (даний рівень визначається на основі експертної оцінки);

- сумарна відстань до маршрутів s_v відстані до яких не перевищують порогового для кластеризації рівня 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 M_s , яка містить елементи $(i, n_s(i), s_v(i))$.

На основі множини M_s , здійснюється формування кластерів та включення до них маршрутів (класифікація маршрутів) за наступною методикою:

1. Якщо для першого елементу множини M_s $n_s(i) > 0$, створюється новий кластер з центральним маршрутом i і до створеного кластеру додаються всі маршрути, відстань до яких від центрального маршруту не перевищує рівня r_{pk} .
2. З множини M_s , видаляється перший елемент i , при виконанні умови $n_s(i) > 0$, всі елементи які відповідають маршрутам доданим до кластера у попередньому кроці.
3. Кроки (1-2) повторюються до тих пір, поки потужність множини M_s є більшою за 0.

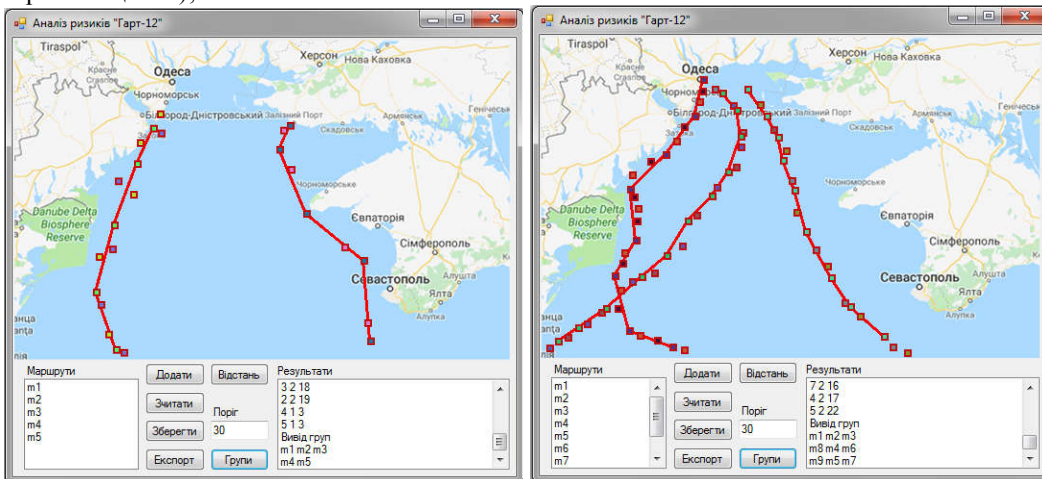


Рис. 4 – Приклади кластеризації даних про маршрути

Описаний вище метод кластеризації маршрутів заданих дискретними відліками в системі висвітлення надводної обстановки отримав алгоритмічну і програмну реалізацію з використанням системи програмування Visual Studio Express Edition та мови програмування С#. З використанням розробленої програми проведена класифікація даних про маршрути, результати якої представлені на рисунку 4 (центральні маршрути кластерів лінійно апроксимовані).

В результаті аналізу наявних даних в ІТС «Гарт-12» з використанням описаного вище методу кластеризації визначаються типові маршрути руху (центри відповідних кластерів). З використанням метрики (6) для нових даних, які поступають до системи висвітлення надводної обстановки, можливо встановити їх належність до одного з стандартних маршрутів та оцінивши ступінь їх подібності, провести оцінку ризиків.

Література

1. Рачок, Р. В. Формування концепції розбудови геоінформаційної системи у сфері забезпечення прикордонної безпеки / Р. В. Рачок // Збірник наукових праць Національної академії Державної прикордонної служби України. Серія : військові та технічні науки / голов. ред. Олексієнко Б. М. – Хмельницький : Вид-во НАДПСУ, 2017. – № 2(72). – С. 295–310. **2. Боровик, О. В.** Визначення підходів до просторового аналізу даних в інформаційно-телекомунікаційній системі морської охорони «Гарт-12» / О. В. Боровик, Р. В. Рачок, В. Ю. Мазур // Збірник наукових праць Національної академії Державної

Висновки й перспективи подальших досліджень

У роботі деталізована методика обчислення метрики для встановлення міри подібності маршрутів суден на основі даних в ІТС морської охорони «Гарт-12». Для забезпечення комутативності цієї метрики, проведено її уточнення.

З метою класифікації маршрутів розроблено метод кластеризації, який дозволяє встановивши належність даних про окремі маршрути до кластерів, визначити їх центри, як маршрути, сумарна відстань від яких до інших маршрутів у кластері є мінімальною.

Проведення кластеризації наявних даних про маршрути дозволяє в подальшому з використанням запропонованої метрики визначити просторові аномалії руху суден та проводити оцінку ступеня ризиків порушення прикордонного законодавства.

Напрямами подальших досліджень є формування методики оцінки ризиків на основі аналізу просторових аномалій руху суден.

прикордонної служби України. Серія : військові та технічні науки / голов. ред. Олексієнко Б. М. – Хмельницький : Вид-во НАДПСУ, 2018. – № 1(75). – С. 295–310. **3. Мандель І. Д.** Кластерный анализ. - М.: Финансы и статистика, 1988. **4. Олдендерфер М. С.,** Блэшфилд Р. К. Кластерный анализ / Факторный, дискриминантный и кластерный анализ: пер. с англ.; Под. ред. И. С. Ен юкова. - М.: Финансы и статистика, 1989. - 215 с. **1. A.K. Jain and R.C. Dubes.** Algorithms for Clustering Data. Prentice Hall, 1988. **2. R. Ng and J. Han.** Efficient and Effective Clustering Method for Spatial Data Mining.

МЕТОД КЛАСТЕРИЗАЦИИ МАРШРУТОВ СУДОВ В СИСТЕМЕ ОСВЕЩЕНИЯ НАДВОДНОЙ ОБСТАНОВКИ

*Валентин Юриевич Мазур (канд. воен. наук, доцент)
Олег Васильевич Боровик (докт. техн. наук, профессор)
Роман Васильевич Рачок (канд. техн. наук, доцент)*

Национальная академия Государственной пограничной службы Украины имени Богдана Хмельницкого, Хмельницкий, Украина

В статье решена задача классификации данных о маршрутах движения судов в системе освещения надводной обстановки. С целью такой классификации предложено использование кластеризации, метод проведения которой разработан в пределах данного исследования. Особенностью метода есть учет специфики представления данных о маршрутах движения судов в информационно-телекоммуникационной системе Морской охраны «Гарт-12». Эта специфика учтена в метрике и особенностях формирования маршрутных кластеров и определение их центров. Такой анализ данных в системе освещения надводной обстановки позволит решать задачу оценки рисков путем выявления пространственных аномалий в маршрутах движения судов.

Описаний в роботі метод кластеризації маршрутів заданих дискретними відліками в системі освітлення надводної обстановки отримав алгоритмічну і програмну реалізацію з використанням системи програмування Visual Studio Express Edition і мови програмування С#. З використанням розробленої програми проведена класифікація даних про маршрути. В результаті аналізу наявних даних в інформаційно-телекомунікаційній системі «Гарт-12» з використанням описаного в роботі методу кластеризації визначаються типові маршрути руху.

движения (центры соответствующих кластеров). С целью классификации маршрутов разработан метод кластеризации, который позволяет определив принадлежность данных об отдельных маршрутах к кластеру можно определить их центры, как маршруты, суммарное расстояние от которых к другим маршрутам в кластере есть минимальным.

С использованием определенной метрики для новых данных, которые поступают к системе освещения надводной обстановки, возможно определить их принадлежность к одному из стандартных маршрутов и проведя оценку степени их сходства провести оценку рисков. Направлениями дальнейших исследований есть формирование методики оценки рисков на основе анализа пространственных аномалий движения судов.

Ключевые слова: информационно-телекоммуникационная система Морской охраны «Гарт-12», система освещения надводной обстановки, кластеризация, метрика, маршрут.

METHOD OF CLUSTERIZATION OF VESSELS ROUTES IN THE SYSTEM OF COVERING THE SITUATION AT SEA

Valentin Y. Mazur (Candidate of Military Sciences, Associate Professor)

Oleh V. Borovyk (Doctor of Technical Sciences, Professor)

Roman V. Rachok (Candidate of Technical Sciences, Associate Professor)

Bogdan Khmelnytsky National Academy of the State Border Service of Ukraine, city of Khmelnytsky, Ukraine

The article solves the problem of classification of data on the routes of the vessels in the system of covering the situation at sea. For the purpose of this classification, the use of clusterization, the method of conducting of which was developed within the framework of this study, is proposed. The peculiarity of the method is to take into account the specificity of presentation the data on the routes of the vessels in the information and telecommunication system of marine guard "Gart-12". This specificity is taken into account in the metric and features of formation of route clusters and definition of their centers. Such an analysis of the data in the system of illumination of the surface environment will solve the problem of risk assessment by identifying spatial anomalies in the routes of the vessels.

Described in the work method of clustering routes specified by discrete reports in the system of illumination of the above-ground situation, received algorithmic and software implementation using a programming system Visual Studio Express Edition and a programming language C#.

Using the developed program, a classification of the route data was carried out. As a result of analysis of the available data in the information system of the «Hart-12» using the clustering method described in the work, typical routes of movement in the (center of the corresponding cluster are determined).

With the aim of classifying routes, a clustering method has been having determined that allows the identified data belonging to individual routes to the cluster to be determined by their centers as routes the total distance from which to the other routes in the cluster is minimal. With the use of a certain metric for new data that comes to the system of lighting the surface situation it is possible to determine their belonging to one of the standard routes and having conducted the assessment of their similarity to carry of their out a risk assessment.

Directions of further research is the formation of a methodology for risk assessment based on the analysis of spatial anomalies of vessel traffic.

Key words: information and telecommunication system of marine guarding "Gart-12", system of covering the situation at sea, clusterization, metric, route.

References

- Rachok, R.V.** (2017) Formation of the concept of building a geographic information system in the field of border security provision [Formuvannya kontseptsiy i rozbudovy heoinformatsiynoyi systemy u sferi zabezpechennya prykordonnoyi bezpeky] / R.V. Rachok // Collection of scientific works of the National Academy of the State Border Guard Service of Ukraine. Series: Military and Technical Sciences / Heads. Ed. Alekseyenko B. M. - Khmelnytsky: NADPSU, 2017. - No. 2 (72). - P. 295-310.
- Borovyk, O.V.** (2018) Definition of approaches to spatial analysis of data in the information and telecommunication system of maritime security "Gart-12" [Vyznachennya pidkhodiv do prostorovoho analizu danykh v informatsiyno-telekomunikatsiyniy systemi mors koyi okhorony «Hart-12»] / O. V. Borovyk, R. V. Rachok, V. Yu. Mazur // Collection of scientific works of the National Academy of the State Border Guard Service of Ukraine. Series: Military and Technical Sciences / Heads. Ed. Alekseyenko B. M. - Khmelnytsky: NADPSU, 2018 - No. 1 (75). - P. 295-310.
- Mandel, I. D.** (1988) Cluster analysis. - Moscow: Finance and Statistics.
- Oldender M. S., Blasfield R. K.** (1989) Cluster Analysis / Factor, Discriminant, and Cluster Analysis: translated by I. S. Yenukova. - Moscow: Finance and Statistics, 215 p.
- A.K. Jain and R.C. Dubes.** Algorithms for Clustering Data. Prentice Hall, 1988.
- R. Ng and J. Han.** Efficient and Effective Clustering Method for Spatial Data Mining.