

В. В. Зосіменко

## Кластеризація як метод оптимального поділу на зіставні групи злочинів із використанням гострих предметів

Київське міське клінічне бюро судово-медичної експертизи

**Ключові слова:** судово-медична експертиза, гострі предмети, кластеризація, таксономічна формула, криміналістична характеристика злочину.

Судово-медична експертиза ушкоджень гострими предметами в судово-медичній практиці посідає особливе місце. Актуальність вивчення можливої наявності певних кореляційних зв'язків між різними елементами криміналістичної характеристики злочину полягає в тому, що дотепер маємо тільки окремі спроби використання для цього саме судово-медичних даних. З метою об'єктивізації ідентифікації злочинця за наявними судово-медичними даними жертви вивчили 849 випадків убивств у Києві, що здійснені з використанням гострого предмета. Результати опрацювали методами варіаційної статистики. Встановили, що кластеризацію можна застосовувати під час аналізу як перший етап симпліфікації даних. Запропонували стислий спосіб опису одержаних класів випадків у вигляді таксономічної формули. Це свідчить про можливість пошуку статистичних зв'язків між показниками злочину та злочинця.

## Кластеризация как метод оптимального разбиения на сопоставимые группы преступлений с применением острых предметов

В. В. Зосименко

Судебно-медицинская экспертиза поврежденных острыми предметами в судебно-медицинской практике занимает особое место. Актуальность изучения возможного наличия корреляционных связей между различными элементами криминалистической характеристики преступления заключается в том, что в настоящее время существуют лишь некоторые попытки использовать для этого именно судебно-медицинские данные. С целью объективизации идентификации преступника по имеющимся судебно-медицинским данным жертвы изучены 849 случаев убийств в Киеве, совершенные с использованием острого предмета. Результаты обработаны методами вариационной статистики. Установлено, что кластеризация может применяться при анализе в качестве первого этапа симплификации данных. Предложен краткий способ описания полученных классов случаев в виде таксономической формулы. Это свидетельствует о возможности поиска статистических связей между показателями преступления и преступника.

**Ключевые слова:** судебно-медицинская экспертиза, острые предметы, кластеризация, таксономическая формула, криминалистическая характеристика преступления.

*Патология.* – 2015. – №2 (34). – С. 106–109

## Clustering as a method of optimal partitioning crimes involving sharp objects into comparable groups

V. V. Zosimenko

Forensic examination of sharps injuries in forensic practice has a special place, since it is connected, as a rule, with crimes against the life and health of citizens. Relevance of the issue for the study of the possible presence of certain correlations between the various elements of criminalistic characteristics of crime is that currently there are only a few attempts to use it for this forensic data.

**Aim.** To identify the objectification of the perpetrator from the available forensic data of a victim.

**Methods and results.** 849 cases of murders with a sharp object in Kiev were studied; results were processed by methods of variation statistics. It was found that clustering can be used in the analysis as a first step simplification of data and offers a very brief way to describe these classes of cases as a taxonomic formula.

**Conclusion.** This indicates that there is an opportunity to find statistical relationships between descriptors (indices) of the crime and the criminal.

**Key words:** Forensic Medicine, Sharp, Cluster Analysis, Criminals

*Pathologia.* 2015; №2 (34): 106–109

Сьогодні маємо усталене визначення терміну «криміналістична характеристика злочину» (КХЗ) – система даних про криміналістично значущі ознаки певного виду або групи злочинів, що характеризується закономірними взаємозв'язками і має на меті створення та перевірку криміналістичних версій [1].

Значення КХЗ полягає передусім у тому, що на початковому етапі розслідування злочинів, коли є певний дефіцит інформації, внаслідок стійких кореляційних зв'язків між різними елементами КХЗ з'являється можливість розгляду нових версій щодо невстановлених обставин. Саме трирівнева структура КХЗ дає можливість

встановити взаємозв'язок між її структурними елементами [2]. Суть трирівневої структури КХЗ, на думку автора, полягає у тому, що елементи КХЗ поділяються на ознаки, а ознаки – на значення. На рівні елементів структура КХЗ включає 4 частини: «Підозрюваний», «Жертва», «Спосіб скоєння злочину», «Місце та час скоєння злочину».

У попередніх роботах навели структуру й ознаки елементів КХЗ, які ми обрали, відзначили наявність і силу статистичного зв'язку між певними показниками [3,4]. Однак механізми об'єктивізації та систематизації статистичних зв'язків між різними показниками КХЗ за наявності різних версій скоєння злочину розроблені недостатньо.

### Мета роботи

Застосування кластеризації на різних рівнях КХЗ, виділення класів (таксонів) та їхній опис для наступного пошуку статистичних зв'язків між дескрипторами злочину та злочинця.

### Матеріали і методи дослідження

Об'єкт дослідження – 849 випадків убивств у Києві (здійснені з використанням гострого предмета) за період 1997–2011 рр., в яких одному злочинцю відповідала одна жертва (труп). Дані одержали з архіву Київського міського клінічного бюро судово-медичної експертизи.

### Результати та їх обговорення

Вирішуючи завдання виявлення характеристик передбачуваного злочинця, на першому етапі необхідно виділити загальні класи злочинів і зіставити їх із класами злочинців. Оскільки класифікація систематизує об'єкти за задалегідь визначеними групами, для побудови оптимального поділу об'єктів на групи зробимо кластеризацію (кластер (від англ. cluster – скупчення) – об'єднання кількох однорідних елементів, яке можна розглядати як самостійну одиницю, що має певні властивості) наявної бази.

Слід розрізнити класифікацію та кластеризацію. Класифікацією називається систематизація кожного елемента за класами із задалегідь відомими параметрами, що отримані на етапі навчання; число класів визначене. Кластеризація – це поділ безлічі даних на кластери, параметри яких задалегідь невідомі; кількість кластерів може бути довільною або фіксованою.

Кластеризацію можна визначити як процес об'єднання даних у групи за схожими ознаками. Оскільки кластеризація розподіляє безліч об'єктів на групи, які визначаються тільки її результатом, оптимальність може бути визначена як вимога мінімізації середньоквадратичної помилки розподілу:

$$e^2(X, L) = \sum_{j=1}^K \sum_{i=1}^{n_j} \|x_i^{(j)} - c_j\|^2,$$

де  $C_j$  – «центр мас» кластера  $j$ .

«Центр мас» кластера – точка у просторі характеристичних векторів із середніми для цього кластера значеннями характеристик. У нашому випадку використовуватимемо кластеризацію для спрощення опрацювання даних і розуміння структури множини об'єктів  $X^i$  за допомогою дивізації (від англ. division – розподіл) на групи схожих об'єктів. Це дасть можливість значно скоротити обсяг збереженої інформації у випадку великої вибірки, залишивши по одному найбільш типовому представнику від кожного кластера. Також можливе виділення нетипових об'єктів, що не відповідають ознакам жодного з кластерів.

У першому випадку число кластерів будемо мінімізувати. У другому важливіше забезпечити високий ступінь подібності об'єктів усередині кожного кластера, а кластерів може бути скільки завгодно. У третьому випадку найбільший інтерес становлять окремі об'єкти, що не відповідають ознакам жодного із кластерів. Для цього наведемо визначення тих понять, якими будемо оперувати. Об'єкт – елементарна група даних, якою оперують

алгоритми кластеризації. Кожному об'єкту відповідає *вектор характеристик*:  $x = (x_1, \dots, x_d)$ . Компоненти  $x_i$  є окремими характеристиками об'єкта. Кількість характеристик  $d$  визначає *розмірність* простору характеристик. Множину, що складається з усіх векторів характеристик, позначатимемо  $X$ :  $X = (x_1, \dots, x_n)$ , де  $x_i = (x_{i1}, \dots, x_{id})$ . *Кластер* – підмножина «близьких один до одного» об'єктів із  $X$ . Відстань  $R(x_p, x_j)$  між об'єктами  $x_i$  та  $x_j$  – результат застосування обраної метрики у просторі характеристик.

У загальному вигляді кластеризація включає етапи виділення характеристик, визначення метрики, поділ об'єктів на групи та наведення результатів (рис. 1).

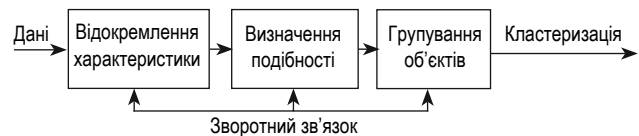


Рис. 1. Алгоритм послідовної кластеризації.

Застосуємо схему виділення кластерів у разі вчинення злочину з використанням гострих предметів. Для виділення вектора характеристик спочатку необхідно обрати властивості, що характеризують об'єкти. Такими можуть бути кількісні (вік злочинця, кількість ран) та якісні ознаки (день тижня, стать, локалізація ран) тощо. Потім потрібно зменшити розмірність простору характеристичних векторів, тобто виділити найбільш важливі властивості об'єктів. Зменшення розмірності прискорює процес кластеризації, а в деяких випадках дає змогу візуально оцінювати її результати. Виділені характеристики потрібно нормалізувати. Далі всі об'єкти представляють у вигляді характеристичних векторів. Ми будемо повністю отожднювати об'єкт із його характеристичним вектором. У всіх цих випадках можна застосовувати ієрархічну кластеризацію, коли більші кластери поділяють на менші, поки в цьому є необхідність.

Такі завдання називають завданнями таксономії (від лат. taxon – порядок, пристрій, організація) – група у класифікації, котра складається з дискретних об'єктів, що об'єднуються на підставі спільних властивостей та ознак і будують не один поділ на класи, які не перетинаються, а систему вкладених поділів. Результат таксономії, як правило, є не простим поділом множини об'єктів на кластери, а являє собою таксономічне дерево – дендрограму, котра дає можливість навести кластерну структуру у вигляді плоского графіка незалежно від того, яка розмірність початкового простору. Замість номера кластера об'єкт характеризується перерахуванням усіх кластерів, яким він належить, від більшого до меншого. Введемо поняття «таксономічної формули» як способу опису кластерної структури, де буквено-цифрові позиції позначають відповідні градації предиктора згідно з таблицями 1, 2.

Формула може бути розширена в разі збільшення розмірності простору (район, пора року тощо) або змен-

Таблиця 1

**Позиція предиктора в таксономічній формулі**

|           |       |     |                |                              |                       |            |                    |
|-----------|-------|-----|----------------|------------------------------|-----------------------|------------|--------------------|
| Позиція   | X     | X   | XXXXX          | Y                            | X                     | X          | X                  |
| Предиктор | Стать | Вік | Зони ушкоджень | Загальна кількість ушкоджень | Місце скоєння злочину | День тижня | Характер дня тижня |
| Приклад   | 1     | 6   | 345            | c                            | 1                     | 7          | 2                  |

Опис злочину: потерпілий – чоловік, вік понад 60 років, локалізація ушкоджень: тулуб, верхні та нижні кінцівки, кількість ушкоджень – від 3 до 10; злочин скоєний у житловому приміщенні в неділю, неробочий день. Таксономічна формула потерпілого: 16345c172.

Таблиця 2

**Градація предиктора в таксономічній формулі**

|                              |  |
|------------------------------|--|
| Предиктор                    | Градація   |
| Стать                        | чоловік – 1; жінка – 2   |
| Вік                          | до 18 – 1; 18–29 – 2; 30–39 – 3; 40–49 – 4; 50–59 – 5; 60 і більше – 6         |
| Зона ушкоджень               | голова – 1; шия – 2; тулуб – 3; верхні кінцівки – 4; нижні кінцівки – 5        |
| Загальна кількість ушкоджень | 1 – A; 2 – B; 3–10 – C; 11–20 – D; 21–50 – E; більше ніж 50 – F                |
| Місце скоєння злочину        | Житлове приміщення – 1; нежитлове приміщення – 2; вулиця – 3; не визначене – 4 |
| День тижня                   | Понеділок – 1; ...; неділя – 7   |
| Характер дня тижня           | Робочий день – 1; вихідний день – 2  |

шена шляхом об'єднання кластерів. Серед алгоритмів ієрархічної кластеризації розрізняють два основних типи. Дивізійні, або низхідні алгоритми розбивають вибірку на все менші кластери. Більш поширеними є агломеративні, або висхідні алгоритми, в яких об'єкти об'єднуються в більші кластери. Реалізація цієї ідеї представлена у такому алгоритмі. Спочатку кожен об'єкт вважають окремим кластером. Для одноелементних кластерів природним чином визначають функцію відстані:  $R(\{x\}, \{x'\}) = p(x, x')$ . Потім запускають процес об'єднання. На кожній ітерації замість пари найближчих кластерів  $U$  і  $V$  утворюється новий кластер  $W = U \cup V$ . Відстань від нового кластера  $W$  до будь-якого іншого кластера  $S$  обчислюється за відстанями  $R(U, V)$ ,  $R(U, S)$ ,

$R(V, S)$ , які до цього моменту мають бути відомі.  $R(U \cup V, S) = \alpha_U R(U, S) + \alpha_V R(V, S) + \beta R(U, V) + \gamma |R(U, S) - R(V, S)|$ , де  $\alpha, \beta, \gamma$  – числові параметри. Ця універсальна формула узагальнює майже всі розумні способи визначити відстань між кластерами (запропонована Вільямсом і Лансом у 1967 р. [5]).

Число кластерів найпростіше визначати шляхом відсічення фінальної ділянки дендрограми. На горизонтальній осі знаходиться інтервал максимальної довжини  $|R_{t+1} - R_t|$ , а як результуюча кластеризація видається множина кластерів  $S_t$ . Число кластерів дорівнює  $K = l - t + 1$ . Якщо є необхідність, можна задати обмеження на мінімальне та максимальне число кластерів  $K_0 < K < K_1$  (табл. 3).

Таблиця 3

**Кластери, що трапляються найчастіше при 7-вимірному просторі характеристик**

| Таксономічна формула | Частота випадків |     | Опис кластера  |
|----------------------|------------------|-----|--|
|                      | Абс.             | %   |  |
| 123a111              | 5                | 0,6 | Чоловік, 18–29 років, 1 ушкодження тулуба, в житловому приміщенні, в понеділок, робочий день |
| 123a172              | 7                | 0,8 | Чоловік, 18–29 років, 1 ушкодження тулуба, в житловому приміщенні, в неділю, неробочий день  |
| 123a341              | 6                | 0,7 | Чоловік, 18–29 років, 1 ушкодження тулуба, на вулиці, в четвер, робочий день                 |
| 123a351              | 7                | 0,8 | Чоловік, 18–29 років, 1 ушкодження тулуба, на вулиці, в п'ятницю, робочий день               |
| 123a362              | 12               | 1,4 | Чоловік, 18–29 років, 1 ушкодження тулуба, в житловому приміщенні, в суботу, неробочий день  |
| 123a372              | 8                | 0,9 | Чоловік, 18–29 років, 1 ушкодження тулуба, на вулиці, в неділю, неробочий день               |
| 133a111              | 9                | 1,1 | Чоловік, 30–39 років, 1 ушкодження тулуба, в житловому приміщенні, в понеділок, робочий день |
| 133a141              | 8                | 0,9 | Чоловік, 30–39 років, 1 ушкодження тулуба, в житловому приміщенні, в четвер, робочий день    |
| 133a162              | 12               | 1,4 | Чоловік, 30–39 років, 1 ушкодження тулуба, в житловому приміщенні, в неділю, неробочий день  |
| 133a172              | 8                | 0,9 | Чоловік, 30–39 років, 1 ушкодження тулуба, в житловому приміщенні, в суботу, неробочий день  |
| 143a111              | 11               | 1,3 | Чоловік, 40–49 років, 1 ушкодження тулуба, в житловому приміщенні, в понеділок, робочий день |
| 143a121              | 7                | 0,8 | Чоловік, 40–49 років, 1 ушкодження тулуба, в житловому приміщенні, у вівторок, робочий день  |
| 143a131              | 7                | 0,8 | Чоловік, 40–49 років, 1 ушкодження тулуба, в житловому приміщенні, в середу, робочий день    |
| 143a172              | 7                | 0,8 | Чоловік, 40–49 років, 1 ушкодження тулуба, в житловому приміщенні, в неділю, неробочий день  |
| 153a121              | 6                | 0,7 | Чоловік, 50–59 років, 1 ушкодження тулуба, в житловому приміщенні, у вівторок, робочий день  |
| 153a162              | 7                | 0,8 | Чоловік, 50–59 років, 1 ушкодження тулуба, в житловому приміщенні, в суботу, неробочий день  |

У нашому випадку інтерес становить не тільки таксономічне дерево загалом, але й визначення оптимального числа кластерів для наступної класифікації, побудови системи класифікаційного зіставлення елементів структури КХЗ.

Дотепер немає універсального алгоритму кластеризації, який був би ефективним на даних різної природи. Переважно використовують ітеративні методи. Водночас результат їхнього застосування суттєво залежить від правильності оцінювання кількості кластерів і розмірності простору характеристик.

## Висновки

1. Кластеризація може бути застосована під час аналізу як перший етап симпліфікації даних. Після виділення схожих груп необхідно застосовувати інші методи.

2. Опис показників трирівневої структури КХЗ (елементи, ознаки, значення) в характеристичних просторах різної розмірності (3-вимірному, 7-вимірному тощо) у вигляді «таксономічної формули» надалі дасть змогу (що і є метою досліджень) мати на початковому етапі розслідування злочинів, коли є певний дефіцит інформації, можливість розгляду версій щодо невстановлених обставин, зокрема визначення показників («таксономічної формули») злочинця.

## Список літератури

1. Балугина Т.С. Криминалистика. Общие положения методики расследования преступлений: лекции / Т.С. Балугина. – Краснодар : КубГУ, 2003.
2. Фесик П.Ю. Технология использования криминалистической характеристики в раскрытии убийств : дис. на соискание ученой степени к.ю.н. : 12.00.09 / П.Ю. Фесик. – Н. Новгород, 2011. – 239 с.
3. Зосіменко В.В. Статистичне моделювання ознак, які індивідуалізують особу за наслідками її агресивних дій з використанням гострих предметів / В.В. Зосіменко // Судово-медична експертиза. – 2014. – №2. – С. 23–27.
4. Зосіменко В.В. Експертний аналіз випадків вбивств із застосуванням гострих предметів як основа криміналістичної характеристики злочину / В.Д. Мішалов, В.В. Зосіменко // Інтегративна антропологія. – 2014. – №2(24). – С. 33–37.
5. Уильямс У.Т. Методы иерархической классификации / У.Т. Уильямс, Д.Н. Ланс // Статистические методы для ЭВМ : пер. с англ. / под ред. М.Б. Малютова. – М. : Наука, 1986. – С. 269–301.

## References

1. Balugina, T. S. (2003). *Kriminalistika. Obschie polozheniya metodiki rassledovaniya prestuplenij* [Criminalistics. General the methods of investigation of crimes]. Krasnodar : KubSU [in Russian].

2. Fesik, P. Yu. (2011). *Tehnologiya ispolzovaniya kriminalisticheskoy kharakteristiki v raskrytii ubijstv* (Dis...kand. med. nauk). [The technology of using criminological characteristics in the disclosure of murders Dr. med. sci. diss.]. Nizhnij Novgorod [in Russian].
3. Zosimenko, V. V. (2014). Statystychni modeliuvannia oznak, yaki indyvidualizuiut osobu za naslidkamy yii ahresyvykh dii z vykorystanniam hostrykh predmetiv [Statistical modeling features which individualize the person on the consequences of its acts of aggression with the use of sharp objects]. *Sudovomedychna ekspertyza*, 2, 23–27 [in Ukrainian].
4. Mishalov, V. D., & Zosimenko, V. V. (2014). Ekspertnyi analiz vbyvstv iz zastosuvanniam hostrykh predmetiv yak osnova kryminalistychnoi kharakterystyky zlochyynu [Expert analysis of homicide cases using sharp objects as the basis of crime characteristic]. *Intehratyvna antropolohiia*, 2(24), 33–37 [in Ukrainian].
5. Williams, W. T., & Lance, G. N. (1986). *Metody ierarkhicheskoy klassifikacii* [Hierarchical classificatory methods]. *Statisticheskie metody dlya EVM*. M.B. Maluytov (Ed.). Moscow: Nauka. [in Russian].

## Відомості про автора:

Зосіменко В.В., зав. відділу комісійних судово-медичних експертиз, Київське міське клінічне бюро судово-медичної експертизи, E-mail: zosimsme@i.ua.

## Сведения об авторе:

Зосіменко В.В., зав. отделом комиссионных судебно-медицинских экспертиз, Киевское городское клиническое бюро судебно-медицинской экспертизы, E-mail: zosimsme@i.ua.

## Information about author:

Zosimenko V.V., Head of the Department Commission Forensic, Kyiv City Clinical Bureau Of Forensic Medical, E-mail: zosimsme@i.ua.

Надійшла в редакцію 17.04.2015 р.