

ОРИГІНАЛЬНІ ДОСЛІДЖЕННЯ

УДК 340.64:343.97:616-035.8

**ГІПОТЕТИКО-ІНДУКТИВНИЙ МЕТОД ЯК ОСНОВА
ІНФОРМАЦІЙНО-НАВЧАЛЬНОЇ ЕКСПЕРТНОЇ СИСТЕМИ У ВИПАДКАХ
СМЕРТІ ВІД ПОРАНЕНЬ ГОСТРИМИ ПРЕДМЕТАМИ**© В. В. Зосіменко¹ А. Ю. Зольнікова²¹Київське міське клінічне бюро судово-медичної експертизи²Національна медична академія післядипломної освіти імені П. Л. Шупика, Київ

Резюме. У статті викладені підходи до побудови інформаційно-навчальної експертної системи класифікації злочинця, як інструменту для аналізу випадків вбивств з використанням гострих предметів. Нами був визначений і перевірений алгоритм послідовної діагностичної процедури, заснований на використанні формули Байєса, як один з можливих серед технологій Data Mining. В подальшому він буде застосований для побудови інформаційно-навчальної експертної системи, заповнюючи яку відомими елементами криміналістичної характеристики злочину, будуть отримані характеристики (стать, вік) злочинця з певною статистичною вірогідністю.

Ключові слова: гострі предмети, інформаційно-навчальна експертна система, статистичний аналіз, криміналістична характеристика злочину, алгоритм послідовної діагностичної процедури.

ВСТУП

Побудова експертної системи класифікації злочинця, як інструменту для аналізу випадків вбивств з використанням гострих предметів, залишається актуальною темою сьогодення. Протягом 2011-2014 років, нами було проведене ретроспективне дослідження 849 неселективних випадків смерті у м. Києві від поранень, які спричинені гострим предметом [1-3]. Зокрема, ми навели обґрунтування того, що на сьогоднішній день для побудови інформаційно-навчальної експертної системи (ІНЕКС) з метою аналізу подібного роду даних, доцільно використовувати технології Data Mining (видобуток даних, вилучення інформації, інтелектуальний аналіз даних).

Метою даної роботи було визначення методом Кульбака найбільш інформативних ознак при вбивствах з використанням гострих предметів (у місті Києві) та побудова таблиці діагностичних коефіцієнтів для визначення класу об'єкта на прикладі ознаки статі злочинця, як однієї зі складових криміналістичної характеристики злочину.

Матеріали та методи дослідження. Об'єктом дослідження були 1068 «Активів судово-медичного дослідження трупів» та «Висновків експерта» з приводу вбивств із використанням гострого предмету, отриманих з архіву Київського міського бюро судово-медичної експертизи. Глибина дослідження складала 15 років (1997-2011 р.р.). Серед них - 930 випадків з наявністю даних про злочинця (вік, стать), з яких нами були обрані 849 випадків, у яких одному злочинцю відповідала одна жертва (труп).

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ ТА ЇХ ОБГОВОРЕННЯ

Великий обсяг інформації, який зберігається в електронному вигляді у базі даних, містить значну кількість прихованих закономірностей, що є досить важливим для прийняття рішень. При цьому, це різномірні дані: бінарні - стать, характер дня і т.п., порядкові - день тижня, кількісні - вік, кількість ран.

Накопичивши достатню кількість прецедентів в базі, можна вирішити різні завдання. Опис злочину за ознаками проводився за допомогою введеної нами таксономічної формули [4], при цьому, збільшення числа атрибутів призводить до збільшення діапазону пошуку. Для скорочення діапазону пошуку та зменшення відсотку помилкових і невизначених відповідей, які можуть бути отримані від ІНЕКС, необхідно використовувати знання про предметну область та евристику.

При вивченні об'єктів, які характеризуються великим числом факторів, найчастіше буває важливо визначити, які з цих факторів більшою мірою впливають на властивості об'єктів, що нас цікавлять. Зокрема, визначення інформативності факторів - це один з важливих етапів аналізу досліджуваного об'єкта [5].

Дані про характеристики злочину зібрані у файл, де кожен рядок являє собою вектор з N значень-ознак. Ознаки нерівнозначні між собою, тому пошук ознак найбільш інформативних для розпізнавання класу об'єкта, є важливою умовою побудови ІНЕКС. Для виявлення найбільш інформативних ознак з подальшою побудовою таблиці діагностичних коефіцієнтів визначення класу об'єкта (статі злочинця), ми застосували метод Кульбака.

Метод Кульбака, при якому обсяги вибірки спостережень за двома класами можуть бути різні, пропонує в якості оцінки інформативності міру розбіжності між двома класами, яка називається дивергенцією, і такий метод, в медицині взагалі, отримав широке застосування при розгляді окремих факторів, які впливають на постановку діагнозу (висновок). В основі методу лежить алгоритм послідовної діагностичної процедури, який впливає з основних теорем теорії ймовірностей і, зокрема, із формули Байєса, яка досить часто застосовується при обчислювальній діагностиці [6].

В основі розроблюваної ІНЕКС лежить гіпотетико-індуктивний метод, теоретичною передумовою якого є доктрина зворотної ймовірності, яка доповнена результатами теорії суб'єктивної ймовірності та статистики, у т.ч. теоремою Байєса як універсальною моделлю пізнання, яка дозволяє апостеріорі вирахувати ймовірності гіпотез з будь-яким не нульовим розподілом апіорних ймовірностей, та теоремою підтвердження Леонарда Севіджа, яка зв'язала теорему Байєса із законом великих чисел [7]. Взагалі, з ім'ям Томаса Байєса (англійського священика і математика XVIII століття), в сучасній статистиці та теорії індукції пов'язано три твердження: **постулат** Байєса – коли допускається рівна апіорна ймовірність гіпотез, **правило поведінки** Байєса – необхідно приймати ту гіпотезу, яка має вищий ступінь апостеріорної ймовірності та **теорема** Байєса – апостеріорна ймовірність гіпотези прямо пропорційна добутку її апіорної ймовірності на її правдоподібність і зворотно пропорційна абсолютної ймовірності її свідчення.

При введенні певних обмежень діагностичної задачі і початкових умов формула Байєса приймає більш простий вигляд (1):

$$\text{поріг } B' < \frac{P(x_1^1/A)}{P(x_1^1/B)} \times \frac{P(x_2^3/A)}{P(x_2^3/B)} \cdots \frac{P(x_j^i/A)}{P(x_j^i/B)} < \text{поріг } A$$

де $\frac{P(x_j^i/A)}{P(x_j^i/B)}$ - відношення правдоподібності для виявленої i -ї градації j -ої ознаки при належності до групи А, до ймовірності цієї ж градації тієї ж ознаки при належності до групи В [8].

Ми повинні заздалегідь вирішити, яке підсумкове перевищення ймовірності належності до класу А над ймовірністю приналежності до класу В (і, навпаки, ймовірності приналежності до класу В над ймовірністю приналежності до класу А) достатньо для прийняття відповідної діагностичної гіпотези.

Відносини ймовірностей, які достатні для прийняття гіпотез А або В, і називають порогами. Таким чином, збір діагностичної інформації і перемноження відносин ймовірностей ознак, які знайдені, продовжують до тих пір, поки правильна вище наведена нерівність, тобто, поки не досягнуть «поріг А» або «поріг В». Коли дана нерівність стає невірним, тобто коли один з порогів досягнутий або перевищений, послідовну процедуру розпізнавання переривають і «виносять рішення» - обирають одну із діагностичних гіпотез. Якщо досягнутий «поріг А», об'єкт відносять до класу А, якщо досягнутий «поріг В» - об'єкт відносять до класу В.

Якщо використана вся наявна у нашому розпорядженні інформація, але жоден з порогів жодного разу не досягнутий, то приймається рішення про те, що «наявної інформації недостатньо для прийняття рішення з визначеним рівнем помилок» (так звана «невизначена відповідь»).

У математичній статистиці подібний підхід, коли послідовне накопичення інформації триває лише до моменту досягнення порогу, запропонував А. Вальд [9].

Він є відмінною особливістю послідовного статистичного аналізу. А. Вальд показав, що при такому підході потрібно в середньому вдвічі менше інформації для прийняття рішення з певним рівнем надійності. Таким чином, при послідовній процедурі розпізнавання для вибору діагностичної гіпотези, ми використовуємо тільки ту частину діагностичної інформації (у вигляді добутку відносин правдоподібності), яка потрібна для досягнення певного порогу. Величина останнього визначається величинами допустимих помилок першого і другого роду. З міркувань зручності обчислень, доцільно добуток відносин правдоподібності замінити відповідним йому складанням логарифмів цих величин. Для того, щоб ці логарифми представляли собою цілі числа, їх множать на 10 і округлюють з точністю до одиниці. Величину, яку при цьому отримують, називають діагностичним коефіцієнтом (ДК). Діагностичний коефіцієнт i градації ознаки x_j дорівнює:

$$ДК(x_j^i) = 10 \lg \frac{P(x_j^i/A)}{P(x_j^i/B)}.$$

Відповідно і діагностичні пороги повинні бути замінені їх логарифмами, помноженими на коефіцієнт 10. «Поріг А» являє собою позитивну величину. При його досягненні приймають гіпотезу А. Поріг В є негативною величиною. При його досягненні приймають гіпотезу В. Замість обчислення порогів можна після прийняття допустимого рівня помилок першого і другого роду визначити їх за допомогою таблиці 1, де α - допустимий відсоток помилок першого роду (коли об'єкт класу А, приймають за об'єкт класу В), а β - допустимий відсоток помилок другого роду (помилкова діагностика класу А, коли насправді об'єкт належить до класу В).

Таблиця 1

Порогові (граничні) суми діагностичних коефіцієнтів

$\alpha \backslash \beta$	20%	10%	5%	2%	1%
20%	+6 -6	+6,5 -9	+7 -12	+7 -16	+7 -19
10%	+9 -6,5	+9,5 -9,5	+10 -12,5	+10 -16,5	+10 -19,5
5%	+12 -7	+12,5 -10	+13 -13	+13 -17	+13 -20
2%	+16 -7	+16,5 -10	+17 -13	+17 -17	+17 -12
1%	+19 -7	+19,5 -10	+20 -13	+20 -17	+20 -20

Формула для розрахунку першого порогу (для визначення рішення – клас А):

$$\text{Поріг А} = 10 \lg \frac{100 - \alpha}{\beta}.$$

Формула для розрахунку другого порогу (для винесення рішення – клас В):

$$\text{Поріг В} = 10 \lg \frac{\alpha}{100 - \beta}.$$

Фактично, вище наведену нерівність (1) можна зобразити інакше, оскільки його складові являють собою діагностичні коефіцієнти відповідних діапазонів відповідних ознак (2):

$$10 \lg \frac{\alpha}{1 - \beta} < \text{ДК}(x_1) + \text{ДК}(x_{21}) + \dots + \text{ДК}(x_j) < 10 \lg \frac{1 - \alpha}{\beta}.$$

Послідовну процедуру визначення величини діагностичних ознак, відшукування відповідних їм діагностичних коефіцієнтів та їх підсумовування продовжують, поки правильна нерівність (2), а коли воно стає невірним (тобто, коли досягнутий або перевищений один з порогів), послідовну процедуру розпізнавання переривають і виносять то чи інше рішення, залежно від того, який з порогів досягнутий. Діагностичні коефіцієнти обчислюють і для всіх інших діапазонів першого і інших ознак, які записують у вигляді діагностичної таблиці (таблиця 2) в порядку інформативності, яка зменшується. При використанні послідовної діагностичної процедури, зручною мірою для оцінки інформативності ознак є міра Кульбака, яка дозволяє оцінити не вірогідність відмінностей між розподілами, а ступінь цих відмінностей [10].

Спочатку обчислюють інформативність градацій діапазонів ознаки. Відповідно до формули Кульбака величина інформативності I діапазону i ознаки j дорівнює (3):

$$I(x_j^i) = \text{ДК}(x_j^i) \frac{1}{2} [P(x_j^i/A) - P(x_j^i/B)].$$

Однак для визначення порядку використання ознак в діагностичній таблиці необхідно обчислити інформативність не одного діапазону ознаки, а всієї ознаки. Інформативність всієї ознаки дорівнює сумі інформативності його діапазонів

$$I(x_j) = \sum I(x_j^i).$$

Після проведення розрахунків ми отримали ознаки з найбільшою інформативністю. У їх число входять: місце злочину, стать жертви, кількість ран. Доцільно розрізнити попередній (можна проводити за допомогою будь-якого критерію відмінностей) і остаточний відбір (проводиться після обчислення інформативності по Кульбаку попередньо відібраних ознак) ознак по інформативності. Питання про поріг інформативності для остаточного відбору ознак вирішується на етапі перевірки діагностичної таблиці по основній і контрольній вибірці з бази даних.

Діагностичні коефіцієнти та інформативність ознак для визначення статі злочинця

Ознака (X_j)	i	Частота спостережень				$DK(X_j^i)$	$I(X_j^i)$	$I(X_j)$	
		Клас 1 (чоловіки)		Клас 2 (жінки)					
		Абс.	%	Абс.	%				
Стать жертви	1	540	77,4	141	93,4	-1,00	0,08	0,95	
	2	158	22,6	10	6,6	5,00	0,87		
Кількість ран	1	280	41,1	82	54,3	-1,00	0,05	0,32	
	2	148	21,1	39	25,8	-1,00	0,03		
	3	157	22,4	20	13,2	2,00	0,02		
	4	59	8,4	5	3,3	4,00	0,10		
	5	42	6,0	4	2,6	4,00	0,07		
	6	12	1,7	1	0,7	4,00	0,04		
Зона ушкодження	Голова	1	133	19,1	15	9,9	3,00	0,05	0,05
		2	565	80,9	136	90,1	0,00	0,00	
	Шия	1	215	30,8	25	16,6	3,00	0,14	0,21
		2	483	69,2	126	83,4	-1,00	0,07	
	Тулуб	1	605	86,7	128	84,8	0,00	0,00	0,02
		2	93	13,3	23	15,2	-1,00	0,02	
	Руки	1	206	29,5	26	17,2	2,00	0,11	0,2
		2	492	70,5	125	82,8	-1,00	0,09	
	Ноги	1	60	8,6	15	9,9	-1,00	0,01	0,01
		2	638	91,4	136	90,1	0,00	0,00	
Місце	1	377	55,2	132	90,4	-2,00	0,33	1,60	
	2	97	14,2	4	2,7	7,00	0,40		
	3	209	30,6	10	6,8	7,00	0,87		
Тип дня	1	407	58,3	98	64,9	0,00	0,00	0,04	
	2	291	41,7	53	35,1	1,00	0,04		
День тижня	1	106	15,2	21	15,9	0,00	0,00	0,15	
	2	81	11,6	16	10,6	0,00	0,00		
	3	72	10,3	22	11,6	-2,00	4,95		
	4	85	12,2	11	9,3	1,00	1,24		
	5	103	11,8	31	20,5	-1,00	1,86		
	6	131	19,2	19	12,6	2,00	7,19		
	7	117	16,8	25	16,6	0,00	0,00		

де: i - номер градації ознаки; $DK(X_j^i)$ - діагностичний коефіцієнт градації ознаки; $I(X_j^i)$ - інформативність градації ознаки; $I(X_j)$ - інформативність ознаки.

Ознаки, відмінності між якими в групах А і В не є статистично значущими, доцільно виключити з подальшої роботи. Однак якщо інформативних ознак мало, а програма дозволяє обробити велику кількість ознак, виявляється доцільним залишити для подальшої роботи і ті ознаки, у яких відмінності між групами А і В недостатньо значимі ($p > 0,05$).

Остаточний висновок про інформативність всього набору ознак можна зробити тільки після перевірки складеної діагностичної таблиці на досить великій перевірочній групі спостережень. Якщо число помилок виявиться велике, можна надалі виключити найменш інформативні ознаки. При цьому, як правило, число помилок зменшується, але число невизначених відповідей зростає.

Для перевірки (випробування) діагностичної таблиці спочатку використовували той матеріал, на якому вона була складена, вважаючи кожен випадок як новий. Перевірку здійснювали за двома вибірками - основної та контрольної. Результати перевірки не розрізнялися ($p > 0,05$) і наведені в таблиці 3.

Таблиця 3

Результати перевірки діагностичної таблиці

Групи	Кількість випадків	Відповіді					
		правильні		помилкові		невизначені	
		Абс.	%	Абс.	%	Абс.	%
Основна	594	439	73,9	73	12,3	82	13,8
Контрольна	255	185	72,5	27	10,6	43	16,9
Загальна	849	624	73,5	100	11,8	125	14,7

Для зменшення кількості помилкових і невизначених відповідей, експерт повинен проводити аналіз отриманих результатів та вивід емпіричних правил або винятків для застосування при побудові ІНЕКС.

Так у випадку, коли було нанесене 1 колото-різане поранення, від якого і настала смерть потерпілого, це не враховувалося при перевірці діагностичної таблиці 3, та вносило помилку (7 % - по основній групі) при визначенні статі злочинця.

Таблиця 4

Результати перевірки діагностичної таблиці з застосуванням емпіричних правил

Групи	Кількість випадків	Відповіді					
		правильні		помилкові		невизначені	
		Абс.	%	Абс.	%	Абс.	%
Основна	594	482	81,1	56	9,4	56	9,4
Контрольна	255	202	79,2	28	10,9	25	9,8
Загальна	849	684	80,6	84	9,9	81	9,5

Як приклад, наведемо визначення суми діагностичних коефіцієнтів з урахуванням їх порогового (граничного) значення, для визначення статі злочинця:

Таксономічна формула потерпілого: 231234f162 (2 - жінка; 3 - вік 30-39 років; 1, 2, 3, 4 - локалізація ран на голові, шиї, тулубі та верхніх кінцівках; f - кількість ран більш ніж 50; 1 - скоєння злочину в житловому приміщенні; 6 - скоєння злочину в суботу; 2 - скоєння злочину не в робочий день). Сума ДК - 13 балів: злочинець - чоловік з імовірністю 95%.

Таксономічна формула потерпілого: 2334c362 (2 - жінка; 3 - вік 30-39 років; 3, 4 - локалізація ран на тулубі та верхніх кінцівках; c - кількість ран в межах 3-10; 3 - скоєння злочину на вулиці; 6 - скоєння злочину в суботу; 2 - скоєння злочину не в робочий день). Сума ДК - 15 балів: злочинець - чоловік з імовірністю 95%.

Таксономічна формула потерпілого: 143a111 (1 - чоловік; 4 - вік 40-99 років; 3 - локалізація ран на тулубі; a - кількість ран - 1; 1 - скоєння злочину в житловому приміщенні; 1 - скоєння злочину в понеділок; 1 - скоєння злочину в робочий день). Сума ДК - 5 балів: злочинець - жінка з імовірністю 75%.

ВИСНОВОК

Все вище викладене може бути підґрунтям для розробки в подальшому гіпотетико-індуктивним методом інформаційно-навчальної експертної системи (ІНЕКС) на мові програмування РНР з інтерфейсом користувача, яка буде, використовуватися в навчальному процесі на кафедрах судової медицини, а в подальшому і мати практичне значення.

Література

1. **Зосіменко В.В.** Аналіз випадків вбивств з використанням гострих предметів у м. Києві як основи для обґрунтування криміналістичної характеристики злочину/В.В. Зосіменко//Судово-медична експертиза. - 2011. - №3. - С. 16-20.
2. **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.
3. **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.
4. **Зосіменко В.В.** Кластеризація, як метод оптимальної розбивки злочинців з використанням гострих предметів на порівнянні групи / В.В. Зосіменко//Патологія. - 2015. - № 2 (34). - С. 106-109.
5. **Загоруйко Н.Г.** Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. - Новосибирск: Изд. Института математики, 1999. - 270 с.
6. **Парин В.В.** Введение в медицинскую кибернетику / В.В. Парин, Р.М. Баевский. - Москва: Медицина, 1966. - 300 с.

7. **Светлов В.А.** Философия и методология науки: учеб. пособие. Ч.1 / В.А. Светлов, И.А. Пфаненштиль. - Красноярск: Сибирский федеральный ун-т, 2011. – С. 356.
8. **Гублер Е.В.** Вычислительные методы анализа и распознавания патологических процессов / Е.В. Гублер. — Л.: Медицина, 1978. — 296 с.
9. **Вальд А.** Последовательный анализ: пер. с англ.; под ред. Б.А. Севастьянова.- Москва: Физматгиз, 1960. – 328 с.
10. **Кульбак С.** Теория информации и статистика: пер. с англ.; под ред. А.Н. Колмогорова. - Москва: Наука, 1967. – 408 с.

ГИПОТЕТИКО-ИНДУКТИВНЫЙ МЕТОД КАК ОСНОВА ИНФОРМАЦИОННО-ОБУЧАЮЩЕЙ ЭКСПЕРТНОЙ СИСТЕМЫ В СЛУЧАЯХ СМЕРТИ ОТ РАНЕНИЙ ОСТРЫМИ ПРЕДМЕТАМИ

Зосименко В.В., Зольникова А.Ю.

Резюме. В статье изложены подходы к построению информационно-учебной экспертной системы классификации преступника, как инструмента для анализа случаев убийств с использованием острых предметов. Нами был установлен и проверен алгоритм последовательной диагностической процедуры, основанный на использовании формулы Байеса, как один из возможных среди технологий Data Mining. В дальнейшем, на основе гипотетико-индуктивного метода, указанный подход будет применен для построения информационно-учебной экспертной системы, заполняя которую известными элементами криминалистической характеристики преступления, будем получать характеристики (пол, возраст) преступника с определенной статистической достоверностью.

Ключевые слова: острые предметы, информационно-учебная экспертная система, статистический анализ, криминалистическая характеристика преступления, алгоритм последовательной диагностической процедуры.

HYPOTHETICAL-INDUCTIVE METHOD AS A BASIS INFORMATION AND TRAINING EXPERT SYSTEM IN CASES OF DEATH FROM INJURY WITH A SHARP OBJECT

Zosimenko V.V., Zolnikova A.Yu.

Resume. The article describes the approaches to the construction of educational information and expert system classification of the offender, as a tool for the analysis of murders with sharp objects. We have installed and tested an algorithm sequential diagnostic procedure based on the use of Bayes' formula, as one of the possible technologies including Data Mining. In the future it will be used for the construction of information and training expert system, which is known for filling elements of criminalistics characterization of crime, will be obtained characteristics (sex, age) of the criminal with a certain statistical confidence. This information and training expert system will be developed in the PHP programming language with the user interface, and will be used in the educational process at the department of forensic medicine, and in the future and make a difference in the practical work of the forensic experts, especially during examinations in cases of violent death with a sharp object.

Keywords: sharp objects, information and training expert system, statistical analysis, criminalistics characterization of crime, algorithms of sequential diagnostic procedure.

УДК 340.64:343.97:616-035.8

МЕТОДИ ОБРОБКИ ДАНИХ ДЛЯ ВИЗНАЧЕННЯ КЛАСУ ЗАЛЕЖНОЇ ЗМІННОЇ ПРИ АНАЛІЗІ ВИПАДКІВ ВБИВСТВ У М. КИЄВІ З ВИКОРИСТАННЯМ ГОСТРИХ ПРЕДМЕТІВ

©В. В. Зосіменко¹, А. Ю. Зольнікова²

¹Київське міське клінічне бюро судово-медичної експертизи

²Національна медична академія післядипломної освіти
імені П. Л. Шупика, Київ

Резюме. Створення експертної системи класифікації злочинця, як інструменту для аналізу випадків вбивств з використанням гострих предметів, залишається актуальною темою сьогодення. Протягом 2011-2014 років, нами було проведене ретроспективне дослідження 849 неселективних випадків смерті у м. Києві від поранень, які спричинені гострим предметом. Зокрема, було наведене обґрунтування того, що на сьогоднішній день для побудови інформаційно-навчальної експертної