

УДК 004.9:502/504(043.3)

С.В. Голуб<sup>1</sup>, І.В. Бурляй<sup>2</sup><sup>1</sup> Черкаський національний університет імені Богдана Хмельницького, Черкаси<sup>2</sup> Академія пожежної безпеки імені Героїв Чорнобиля, Черкаси

## ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ КЛАСТЕРИЗАЦІЇ ЗА РЕЗУЛЬТАТАМИ МОДЕЛЮВАННЯ В ІНФОРМАЦІЙНІЙ ТЕХНОЛОГІЇ ОПЕРАТИВНОГО ПОЖЕЖНОГО МОНІТОРИНГУ

З метою підвищення ефективності прийняття рішень щодо профілізації підготовки пожежно-рятувальних підрозділів проведено аналіз методів кластеризації масивів вхідних даних (МВД) з метою створення методів формування МВД, що забезпечують підвищення якості результатів моделювання об'єктів та процесів пожежогасіння. Отримано експериментальне підтвердження гіпотези про можливість підвищення ефективності кластеризації за результатами моделювання шляхом уточнення складу кластера. Отримані послідовності спостережень, яким властиві спільні механізми впливовості факторів та дозволяють синтезувати точні, адекватні та структурно стійкі моделі.

**Ключові слова:** кластеризація, масив вхідних даних, матриця, кластер, оптимізація кластеру.

### Вступ

При прийнятті рішень з ліквідації пожежі керівник гасіння пожежі (КПП) враховує обстановку, яка складається на місці виклику і здійснює вибір рішення на базі теоретичних знань набутих за програмою спеціальної підготовки пожежного та практичних навичок, умінь та знань набутих під час здійснення професійної діяльності [1].

Теоретичні знання пожежники здобувають в декілька етапів [2, 3]:

1. Базова професійна підготовка.
2. Службова професійна підготовка.

Базова підготовка отримується в центрах професійної підготовки, вищих відомчих навчальних закладах для осіб, які вперше прийняті на службу до пожежно-рятувальних підрозділів. Службова підготовка проводиться постійно з особовим складом пожежно-рятувальних підрозділів шляхом теоретичних і практичних занять, індивідуального навчання за місцем служби.

Службова підготовка повинна враховувати особливості підготовки пожежників під конкретні умови діяльності зважаючи на те, що базова підготовка є загальною. Однак, організацію службової підготовки формалізовано в нормативних документах [2, 3], які не дозволяють в повному обсязі проводити підготовку особового складу з врахуванням необхідної уніфікації. У вказаних документах приведено примірну навчальну програму проведення занять із службової підготовки з середнім і старшим начальницьким складом органів та підрозділів цивільного захисту Державної служби України з надзвичайних ситуацій (ДСНС).

Дана навчальна програма є уніфікованою для підрозділів і не передбачає внесення достатніх

змін для забезпечення ефективної уніфікованої підготовки.

Таким чином, постає проблема відсутності інформаційної технології у складі системи багаторівневого моніторингу пожежної безпеки, яка покликана розв'язувати актуальну практичну задачу профілізації підготовки рятувальників відповідно до оперативного-тактичних характеристик районів обслуговування пожежно-рятувальних підрозділів.

**Аналіз останніх досліджень і публікацій.** Основне завдання кластерного аналізу – розбиття заданої вибірки об'єктів на підмножини, що називаються кластерами, так, щоб кожен кластер складався зі схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися [4]. Завдання кластеризації відноситься до статистичної обробки, а також до широкого класу завдань навчання без вчителя.

Основна мета кластерного аналізу – знаходження груп схожих об'єктів у вибірці досягається за рахунок оцінювання близькості двох об'єктів один від одного.

Міри близькості обирають виходячи з властивостей об'єктів. Основними є такі множини метрик [5 – 7]:

1. Квадрат евклідової відстані. Застосовується для надання більшої ваги більш віддаленим один від одного об'єктам. Ця відстань обчислюється таким чином:

$$\rho(x, x') = \sum_i^n (x_i - x'_i)^2.$$

2. Відстань міських кварталів (манхетенська відстань). Є середньою різницею по координатам. В більшості випадків ця міра відстані призводить до таких результатів, як і Евклідова відстань. Однак для цієї міри вплив окремих великих різниць (вики-

дів) зменшується –  $\rho(x, x') = \sum_i^n |x_i - x'_i|$ .

3. Степенева відстань. Застосовується у випадках, коли необхідно збільшити чи зменшити вагу, яка відноситься до розмірності, для якої об'єкти суттєво відрізняються. Розраховується за формулою:

$$\rho(x, x') = \sqrt[r]{\sum_i^n (x_i - x'_i)^p},$$

де  $r$  і  $p$  – параметри, визначені користувачем.

Параметр  $r$  відповідальний за поступове зважування різниць по окремим координатам, параметр  $p$  відповідальний за прогресивне зважування великих відстаней між об'єктами. Якщо  $r$  і  $p$  рівні двом, то ця відстань співпадає з відстанню Евкліда.

4. Відстань Журавлова [7]. Формула для визначення відстані:

$$\rho_{ik} = \sum_{j=1}^N I_{ik}^j, \text{ де } I_{ik}^j = \begin{cases} 1, \text{ якщо } |x_{ij} - x_{kj}| < \varepsilon \\ 0, \text{ інакше.} \end{cases}$$

В даній формулі:  $i$  – номер рядка в матриці вихідних даних;  $k$  – номер стовпчика в тій же матриці,  $j$  – послідовний номер стовпчика в оброблюваних рядках;  $x_{ij}$  – елемент рядка  $i$  вихідної матриці;  $x_{kj}$  – елемент рядка  $k$  вихідної матриці;  $\varepsilon$  – додатне як завгодно мале речове число. В даному випадку відстані між  $x_{ij}$  та  $x_{kj}$  в будь-якому випадку додатні. Відповідно можна припустити, що  $\varepsilon$  це деякий критерій мінімальної відстані між точками.

Міра близькості підбирається індивідуально для конкретних типів даних. Інколи адекватної міри близькості підібрати не вдається, і доводиться обирати її евристично.

Далі обирають алгоритм, за яким будують модель даних, тобто групують об'єкти. Вибір алгоритму складний, інколи доводиться використовувати декілька алгоритмів, перш ніж буде отримано потрібний (інтерпретувемий) результат. Алгоритми кластеризації можуть комбінуватися з метою отримання метаалгоритму, результат виконання одного з комбінантів слугує проміжним результатом виконання іншого. Реалізація алгоритму призводить до побудови моделі даних, тобто до групування об'єктів по кластерам.

Інтерпретація результатів проведення кластерного аналізу дозволяє отримати знання, тобто корисні правила, які можна використовувати в подальшому для віднесення нових об'єктів до тієї чи іншої групи – кластеру.

В процесі розв'язання задачі створення методів формування МВД, що забезпечують підвищення якості результатів моделювання об'єктів та процесів пожежогасіння [8] була сформульована гіпотеза, що

кластеризацію спостережень МВД необхідно проводити за результатами їх моделювання. Створений алгоритм кластеризації (рис. 1) використаний під час модельного експерименту підтвердив гіпотезу про доцільність кластеризації спостережень масиву вхідних даних за результатами їх моделювання. Отримані послідовності спостережень, що характеризуються спільними механізмами впливовості факторів та дозволили синтезувати точні, адекватні та структурно стійкі моделі. Похибка моделювання при цьому зменшилася в середньому на 96% [8, 9].



Рис. 1. Алгоритм кластеризації

### Постановка задачі підвищення ефективності кластеризації

В роботі [10, 11] розглянуті проблеми синтезу якісних моделей для забезпечення інформацією процесу оперативного керування пожежогасінням за допомогою методу багаточарового синтезу цих моделей, який здатний забезпечити їхню адекватність в умовах недостатньої інформативності МВД. Продовженням дослідження проблематики використання нових інформаційних технологій з метою підвищення ефективності прийняття управлінських

рішень керівника гасіння пожежі (КГП) шляхом проведення профілізації підготовки пожежних підрозділів та координації робіт з гасіння великих пожеж стала розробка методів та засобів формування масивів вхідних даних та синтезу якісних моделей з проведенням кластеризації спостереження масиву вхідних даних за результатами їх моделювання [8].

Були отримані результати середньоквадратичного відхилення моделювання що не перевищувало 10%.

Формальна задача кластеризації спостережень процесу пожежегасіння. Існує масив вхідних даних у вигляді матриці (1), отриманий як результат спостережень за об'єктами та процесами пожежегасіння

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} & y_{11} & y_{12} & \dots & y_{1m} \\ x_{21} & x_{22} & \dots & x_{2n} & y_{21} & y_{22} & \dots & y_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{k1} & x_{12} & \dots & x_{kn} & y_{k1} & y_{k2} & \dots & y_{km} \end{pmatrix} \quad (1)$$

де  $x_{ij}$  –  $j$ -та характеристика стану підрозділів та об'єкта  $i$ -го спостереження процесу пожежегасіння,  $y_{ij}$  –  $j$ -та характеристика втрат ресурсів  $i$ -го спостереження процесу пожежегасіння,  $k$  – кількість спостережень;  $n$  – кількість характеристик процесу  $i$  підрозділів пожежегасіння та його результатів,  $m$  – кількість характеристик втрат ресурсів та інших результатів пожежегасіння.

Прикладом такої множини може бути вибірка з масиву карток обліку пожежі (табл. 1 [12]).

Кожний рядок матриці відображає окрему точку спостереження за об'єктом у багатовимірному просторі характеристик процесу пожежегасіння. Наперед відомо, що існує перелік спостережень, що поєднують в собі чисельні характеристики показників  $X = \{x_1, x_2, \dots, x_n\}$ , на результати яких  $Y = \{y_1, y_2, \dots, y_m\}$  значимо впливали одні і ті ж фактори

$$Y = f(X, W), \sum_{j=1}^n w_j = 1, \quad (2)$$

Таблиця 1

Характеристики процесу пожежегасіння

Параметр пожежегасіння	Змінна	№ спостереження									
		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
День виклику	$x_1$	30	30	31	31	31	31	31	31	31	31
Місяць виклику	$x_2$	12	12	12	12	12	12	12	12	12	12
Рік виклику	$x_3$	2009	2009	2009	2009	2009	2009	2009	2009	2009	2009
Поверховість будівлі	$x_4$	1		1	1	1	2		1	1	5
Поверх, на якому виникла пожежа	$x_5$	1		1	1	1	2		1	1	4
Ступінь вогнестійкості	$x_6$			3	3	3	3		3	3	2
Час виникнення	$x_7$	0,74	0,19	0,10	0,69	0,86	0,50	0,95	0,95	0,36	0,03
Час повідомлення	$x_8$	0,75	0,19	0,10	0,72	0,88	0,51	0,95	0,96	0,38	0,03
Час прибуття 1-го підрозділу	$x_9$	0,76	0,20	0,11	0,73	0,89	0,52	0,96	0,97	0,38	0,03
Час локалізації	$x_{10}$	0,77	0,20	0,13	0,73	0,90	0,52	0,97	0,99	0,40	0,04
Час ліквідації	$x_{11}$	0,78	0,21	0,20	0,74	0,93	0,54	0,97	1,09	0,41	0,04
Умови, що сприяли поширенню	$x_{12}$				11	11	11			11	13
Умови, що ускладнювали гасіння	$x_{13}$				13	13	13			13	
Учасники гасіння пожежі	$x_{14}$	1	1	1	1	1	1	1	1	2	1
Кількість учасників гасіння пожежі	$x_{15}$	3	5	4	5	4	9	4	4	4	5
Техніка	$x_{16}$	44	11	44	11	11	11	11	11	11	11
Кількість техніки	$x_{17}$	1	1	2	1	1	1	1	1	1	1
Стволи	$x_{18}$	2	2	2	2	2	2	2	2	2	2
Кількість стволів	$x_{19}$	1	1	2	1	1	1	1	1	1	1
Вогнегасні речовини	$x_{20}$	11	11	11	11	11	11	11	11	11	11
Первинні засоби пожежегасіння	$x_{21}$	22	22	22	22	22	22	22	22	22	11
Джерела водопостачання	$x_{22}$	6	6	6	6	6	6	6	6	6	2
Керівник гасіння пожежі	$x_{23}$	11	16	16	11	16	16	11	11	12	14

де  $W = \{w_1, w_2, \dots, w_n\}$  – вектор вагових коефіцієнтів моделі (оцінок впливовості кожного з факторів  $x_i$ , що ввійшли до структури цієї моделі). Оцінки впливовості факторів отримують в результаті дослідження моделей на чутливість, при цьому результати випробування моделей повинні довести їх точність, адекватність та стійкість.

Необхідно для модельованого значення втрат (часу гасіння пожежі) створити метод кластеризації точок спостережень, що дозволяє отримати масив вхідних даних у вигляді матриці (3):

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} & y_1 \\ \dots & \dots & \dots & \dots & y_2 \\ \dots & \dots & \dots & \dots & \dots \\ x_{k1} & x_{k2} & \dots & x_{kn} & y_k \end{pmatrix}, \quad (3)$$

що вміщує послідовність точок спостережень ( $X_i, y_j$ ),  $X_i \in X, i=1, k$ , де  $X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$  – масив характеристик стану підрозділів об'єкту  $i$ -ї точки спостереження процесу пожежогасіння,  $y_j$  –  $j$ -та характеристика втрат ресурсів в результаті пожежі для даної точки спостереження, який дозволяє розв'язати задачу ідентифікації функціональної залежності  $y_j = f(X)$ , за умови одержання максимальної якості моделей синтезованих для кожного кластеру окремо відповідно критерію [8]:

$$\delta = \frac{\sum_{i=1}^k \left( \frac{|y_i - y_i^*|}{y_i} \right)}{n} \rightarrow \min, \quad (4)$$

де  $k$  – кількість спостережень послідовності;  $y_i^*$  – змодельоване значення залежної змінної;  $y_i$  – дійсне значення залежної змінної.

### Постановка задачі підвищення ефективності кластеризації

Існує перелік точок спостереження у вигляді кластера (3), який використано у вигляді масиву вхідних даних (МВД). За цим МВД синтезована модель, за допомогою якої розв'язана задача функціональної ідентифікації (2). За умови, коли значення показника якості (4) не задовольняє дослідника  $S_y > S_y^{доп.}$ , необхідно сформувати такий МВД на базі кластера (3), який би забезпечив необхідну адекватність результатам моделювання:

$$S_y \leq S_y^{доп.} \quad (5)$$

Задача розв'язується шляхом уточнення переліку точок кластеру (3) в результаті їх повторного моделювання та виділення кількох кластерів на його основі відповідно схемі, поданій на рис. 2.

Масив вхідних даних містив 14028 спостережень, що відображали характеристики пожежогасіння на території України впродовж 2009 року.

Ефективність кластеризації точок спостереження зростає, якщо знижується значення критерію (5).

### Результати досліджень

В табл. 2 подані результати дослідження складу кластеру за результатами його повторного моделювання на кожному кроці.

Зниження показника похибки в кроках 1-6 та досягнення мінімуму в кроці 7 свідчить про завершення процесу уточнення складу кластеру (рис. 3).

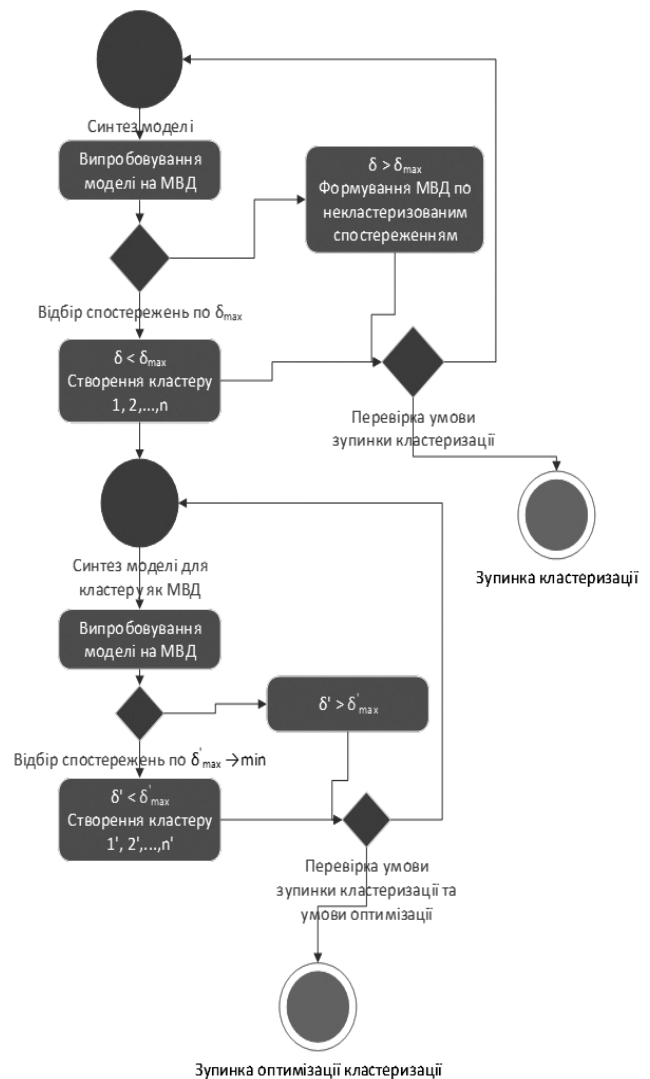


Рис. 2. Діаграма процесу оптимізації складу кластеру

Таблиця 2

Результати моделювання

Номер кроку	Кількість спостережень	СКВ	Середня похибка
1	2362	88,26	5,11%
2	2252	86,29	4,64%
3	2192	85,36	4,48%
4	2154	85,28	4,37%
5	2128	85,07	4,30%
6	2113	84,16	4,28%
7	2103	84,16	4,28%

## Висновки

Отримано експериментальне підтвердження можливості підвищення ефективності процесу кластеризації точок спостереження за результатами їх моделювання шляхом введення додаткового етапу відбору і синтезу нових моделей. Підвищення якості моделі-кластеризатора відбувається в результаті уточнення складу кластеру. Похибка моделювання при цьому зменшується в середньому на 16,2%.

Предметом майбутніх досліджень є виявлення особливостей формування глобальної функції перетворення інформації в технології оперативного пожежного моніторингу при застосуванні процесів підвищення ефективності кластеризації точок спостереження, які утворюють масиви вхідних даних.

## Список літератури

1. Терещев В.В. Пожарная тактика / В.В. Терещев, А.В. Подгрушный. – М., 2007. – 577 с.
2. Положення про організацію службової підготовки осіб рядового і начальницького складу органів та підрозділів цивільного захисту // Наказ МНС України від 01.09.2009 № 601.
3. Про організацію службової підготовки осіб рядового і начальницького складу органів та підрозділів цивільного захисту ДСНС України у 2013/2014 навчально-му році // Наказ ДСНС України від 20.08.2013 № 543.
4. Кластерний аналіз [Електронний ресурс] / Матеріал з Вікіпедії – вільної енциклопедії. – Режим доступу до ресурсу: URL:[http://uk.wikipedia.org/wiki/Кластерний\\_аналіз](http://uk.wikipedia.org/wiki/Кластерний_аналіз) – 24.06.2013 г. - Загол. з екрану.
5. Обзор алгоритмов кластеризации данных [Електронний ресурс] / Хабрахабр. – Режим доступу до ресурсу: URL: <http://habrahabr.ru/post/101338/> – 24.06.2013 г. - Загл. с экрана.

6. Самообучение в интеллектуальных системах. Постановка задачи кластер-анализа. Критерии и метрики кластер-анализа [Електронний ресурс] / Факультет другої вищої та післядипломної освіти ННУ "ІПСА" НТУУ "КПІ". – Режим доступу до ресурсу: URL: [http://iasa.org.ua/lections/tpr/studying/self\\_clust.htm](http://iasa.org.ua/lections/tpr/studying/self_clust.htm) – 24.06.2013 г. – Загл. с экрана.

7. Расстояние Журавлева [Електронний ресурс] / ©2010 Кластерный анализ и поисковые системы. – Режим доступу до ресурсу: [www/ URL: http://blog.iripivf.com/2010/02/05/rasстояние-журавлева/](http://www.blog.iripivf.com/2010/02/05/rasстояние-журавлева/) – 25.06.2013 г. - Загл. с экрана.

8. Голуб С.В. Структуризация массивов входных данных в информационной технологии оперативного мониторингу пожежозасіння / С.В. Голуб, І.В. Бурляй // Радиоелектронні і комп'ютерні системи. – 2013. – № 5 (64). – С. 23-30.

9. Holub S. Classification of observations for technologies of processing of monitoring results of fire fighting process / S. Holub, I. Burlai // ISC UniTech'13, V.II, 2013. – P. 155-161.

10. Багатошарові моделі в технології моніторингу пожежної безпеки з багаторівневим перетворенням інформації [Текст]: звіт про НДР (закл.) 30.11.13 / Академія пожежної безпеки імені Героїв Чорнобиля; керівн. С.В. Голуб; відпов. викон.: І.В. Бурляй. – Черкаси, 2013. – 27 с. – Інв. № 0112U008407.

11. Holub S. Multilayer models with multilevel information transformation in technology of fire safety monitoring / S. Holub, I. Burlai // ISC UniTech'12, V.I, 2012. – P. 464-466.

12. НАПБ 07.017-2004 – Картка обліку пожежі [Текст]. – Затверджено Наказом МНС України від 29.01.04 № 39. – К., 2004.

Надійшла до редколегії 18.02.2014

Рецензент: д-р техн. наук, проф. В.М. Рудницький, Черкаський державний технологічний університет, Черкаси.

## ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ КЛАСТЕРИЗАЦИИ ПО РЕЗУЛЬТАТАМ МОДЕЛИРОВАНИЯ В ИНФОРМАЦИОННОЙ ТЕХНОЛОГИИ ОПЕРАТИВНОГО ПОЖАРНОГО МОНИТОРИНГА

С.В. Голуб, И.В. Бурляй

С целью повышения эффективности принятия решений по профилированию подготовки пожарно-спасательных подразделений проведен анализ методов кластеризации массивов входных данных (МВД) с целью создания методов формирования МВД, обеспечивающих повышение качества результатов моделирования объектов и процессов пожаротушения. Получено экспериментальное подтверждение гипотезы о возможности повышения эффективности кластеризации по результатам моделирования путем уточнения состава кластера. Получены последовательности наблюдений, которым свойственны общие механизмы влияния факторов и позволяют синтезировать точные, адекватные и структурно устойчивые модели.

**Ключевые слова:** кластеризация, массив входных данных, матрица, кластер, оптимизация кластера.

## IMPROVING THE EFFICIENCY OF CLUSTERING BY THE SIMULATION RESULTS IN TECHNOLOGY OF FIRE SAFETY MONITORING

S.V. Holub, I.V. Burlai

Analysis of methods of input data array (IDA) clustering was done aimed to create methods of IDA formation, which could improve quality of simulation of objects and processes of firefighting. Experimental confirmation was received of hypothesis about the possibility of increasing the efficiency of clustering of observations of IDA by refining the composition of the cluster.

**Keywords:** clustering input data array, matrix, cluster, a cluster optimization.