

Нгуен ВьонгАнь  
 Академия Государственной Противопожарной Службы МЧС России  
 129366, г. Москва, ул. Бориса Галушкина, д.4  
 E-mail: vuonganh2709@gmail.com

### Использование метода кластерного анализа для базы данных пожаров во Вьетнаме

В настоящее время во Вьетнаме существует база данных о пожарах, которая служит для статистики и анализа пожарной ситуации. База содержит информацию о муниципальном образовании (численность населения, площадь) о количестве пожаров за текущий год, количестве жертв, количестве профилактических мероприятий и т.д. Статистический анализ на этой базе с применением кластерных и корреляционных методов позволяет с достаточной надежностью сделать выводы об эффективности профилактических мероприятий, применительно к различным муниципальным образованиям, в зависимости от их территории, количества населения и иных факторов. На основании кластерного и корреляционного анализа возможно выделение существенных переменных, влияющих на возможность возникновения пожаров, количество жертв и др., которые в дальнейшем могут быть использованы в качестве входных переменных для обучения систем искусственного интеллекта при прогнозировании пожаров и их последствий.

Кластерный анализ в предлагаемой работе реализовывался методом К – средних в  $d$  – мерном пространстве [1,2].

Коэффициент корреляции вычислялся стандартным образом [1-3]:

$$\text{corr}(x_i, x_j) = \frac{\sum_k^n (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_k^n (x_{ik} - \bar{x}_i)^2 (x_{jk} - \bar{x}_j)^2}},$$

где  $i=1 \dots d$ - размерность пространства;

$x_{ik}$  – реализации случайного процесса;

$k=1 \dots n, \bar{x}_i$  – среднее значение компоненты  $i$  случайной величины  $x_i$ .

Значение  $\text{corr}(x_i, x_j)$  сравнивалось с критическим при заданном уровне риска  $\alpha$  и числе степеней свободы  $r_{cr}(\alpha, n - 1)$ , при  $|\text{corr}(x_i, x_j)| > r_{cr}(\alpha, n - 1)$  корреляция считалась значимой.

В исследовании, проведен анализ базы данных пожаров в Хошимине за трехлетний период 2017-2019 гг, использовались два метода- метод кластерного анализа (алгоритм К- средних) и методы корреляционного анализа. Все данные нормировались на максимальное значения в соответствующей проекции. На рис.1 представлены результаты трехмерного кластерного анализа в пространстве население-пожары –гибель по 13 районам в Хошимине за 2019 год.

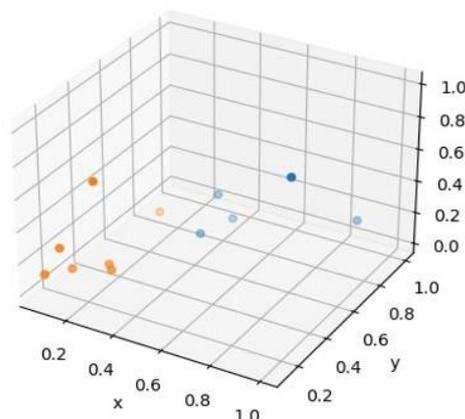


Рис.1 Результаты кластерного анализа в пространстве: население (x) –пожары (y) – гибель (z).

Выделено два кластера - один тяготеет к территориям с небольшим населением и более сгруппирован, другой рассредоточен в 3D пространстве и в целом соответствует более крупным образованиям в Хошимине. В каждом сечении был проведен корреляционный анализ результаты которого представлены в табл.1. В сечениях (население-пожары) и (пожары-гибель) установлены значимые положительные коэффициенты корреляции для обоих кластеров для первого кластера корреляции статистически значимы, с уровнем риска принятия гипотезы 5%, а для второго кластера в сечении население-пожары корреляции значимы с уровнем риска 2%, а пожары – гибель не значима с уровнем риска 5%. Так же отсутствует значимая корреляция в сечении население-гибель, что вполне понятно, учитывая сильную дисперсию населения муниципальных образований. Схожие результаты наблюдались при КА пожаров в российских регионах [4].

Табл.1 Кластерный анализ в различных сечениях;  $r$  - коэффициент корреляции,  $n$  – количество точки в кластере

	Кластер 1		Кластер 2	
	$r$	$n$	$r$	$n$
(xy)	0.673	7	0.893	5
(xz)	0.193	7	0.069	5
(yz)	0.388	7	0.364	5

Таким образом, предварительный кластерный и статистический анализ информации о пожарах в Хошимине за 2017-2019 гг. позволил выделить значимые параметры для профилактической работы и дальнейшего построения моделей или обучения систем искусственного интеллекта при прогнозировании пожаров и их последствий.

#### Литература

1. Воронцов К.В. Алгоритмы кластеризации и многомерного шкалирования. Курс лекций. МГУ, 2007.
2. Jain A., Murty M., Flynn P. Data Clustering: A Review. // ACM Computing Surveys. 1999. Vol. 31, no. 3, p.12.
3. Гмурман В.Е. Теория вероятностей и математическая статистика. Учеб. пособие для вузов. М., «Высшая школа», 1977, с.435
4. Авдеенко А.М., Лахвицкий Г.Н., Нгуен Вьонг Ань, Сатин А.П., Бурлаченко К.Г. Некоторые особенности исследования многомерной базы пожаров с использованием инструментария кластерного анализа Пожарная и аварийная безопасность 17 Международная научно-практическая конференция Иваново 24,11,2022 г.