

В.А. Минаев¹, Н.Г. Топольский², Дао Ань Туан² (Россия, Вьетнам)
(¹МГТУ им. Н.Э. Баумана, ²Академия ГПС МЧС России; e-mail: m1va@yandex.ru)

ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ЗАДАЧИ ТИПОЛОГИЗАЦИИ ТЕРРИТОРИЙ ВЬЕТНАМА ПО ПОЖАРНОЙ ОБСТАНОВКЕ

Изложена постановка задачи типологизации территорий Вьетнама на основе методов кластерного анализа по характеристикам пожарной обстановки в жилом секторе и хозяйствующих субъектах. Произведён отбор значимых показателей для решения задачи типологизации.

Ключевые слова: типологизация, пожарная обстановка, жилой сектор, хозяйствующие субъекты.

V.A. Minaev, N.G. Topolsky, Dao Anh Tuan (Russia, Vietnam)

INFORMATION SUPPORT OF PROBLEM OF VIETNAM'S TERRITORIES FIRE SITUATION TYPOLOGY

Statement of the problems of Vietnam's territories typology on the basis of cluster analysis methods by characteristics of the fire situation in the residential sector and at the business objects sector are given. Selection of significant indicators for solving the problem of typology was carried out.

Key words: typology, fire safety, fire situation, residential sector, business objects sector.

Статья поступила в редакцию Интернет-журнала 27 ноября 2016 г.

Введение

Чтобы учесть территориальные аспекты пожарной обстановки в жилых массивах, промышленных и сельскохозяйственных объектах Вьетнама, необходимо рассмотреть социально-экономические, демографические, организационно-управленческие, материально-технические и другие факторы. Поэтому типологизация территорий по пожарной обстановке должна учитывать объёмы и класс горючих материалов, находящихся на них [1], особенности климата и рельефа местности [2], частоту возникновения источников возгорания [3], наличие и качество противопожарных средств [4] и т.д.

Необходимо отметить, что развитие технологий дистанционного зондирования в последнее время сделало возможным построение детальных и пространственно более полных характеристик пожарной обстановки на территориях [5].

Среди методов многомерной классификации, предназначенных для разделения провинций Вьетнама на однородные группы (типы) в смысле схожести условий по пожарной обстановке, важное место занимает кластерный анализ.

Задача типологизации территорий по пожарной обстановке

Пусть X – множество территорий (провинций) Вьетнама, Y – множество кластеров (типов территорий по пожарной обстановке). Задана некоторая функция расстояния между территориями $\rho(x_i, x_j)$, где i, j – индексы территорий; $i = 1, \dots, n$; $j = 1, \dots, n$; n – общее количество территорий во множестве $X = \{x_i\}$.

Требуется разбить множество территорий X на непересекающиеся подмножества $Y^1 \cup Y^2 \cup \dots \cup Y^M = X$, называемые *кластерами* Y^m , ($m = 1, \dots, M$), так, чтобы каждый кластер (тип территорий) состоял из объектов, близких по метрике ρ , а объекты; разных кластеров существенно отличались по той же метрике. При этом каждому объекту $x_i \in X$ приписывается ещё и номер кластера x_{im} ; $m = 1, \dots, M$.

Алгоритм кластеризации – это функция $\Psi(X) \rightarrow Y$, которая любому объекту $x_i \in X$ ставит в соответствие номер кластера Y^m . Цель реализации алгоритма – определить оптимальное число кластеров, с точки зрения некоторого *критерия качества* кластеризации, отражающего многомерное представление пожарной обстановки для территории страны.

Решение задачи кластеризации, как правило, неоднозначно, что обусловлено тремя основными причинами:

- критерий качества кластеризации является эвристическим, завися от представлений экспертов о предметной области, оценивающих разумность выделения кластеров;
- число кластеров в общем случае априори неизвестно, устанавливаясь в соответствии с некоторым субъективным критерием меры близости оцениваемых объектов;
- результат кластеризации существенно зависит от метрики, выбор которой определяется экспертами.

Задача типологизации является своеобразной "свёрткой" исходных информационные таблиц. При этом число выделяемых типов объектов или явлений всегда меньше, чем уникальных единиц.

В результате получается лаконичное, логичное и наглядное представление типов территорий по пожарной обстановке в пространстве существенно меньшей размерности, позволяющее более точно обосновывать и принимать решения о предупреждении и эффективном тушении пожаров в зависимости от типа территории.

Задача кластерного анализа состоит в изучении по эмпирическим данным, каким образом территории "связываются" в "скопления" – кластеры, при этом никаких априорных предположений о структуре и количестве типов не производится. Таким образом, решается задача разбиения на типы с целью выделения групп однородных территорий, сходных между собой по характеристикам пожарной обстановки, при существенном отличии этих групп друг от друга.

Большинство методов кластеризации [6] основывается на анализе квадратной и симметричной относительно главной диагонали матрицы коэффициентов сходства (расстояния, корреляции и т.д.) между объектами исходной матрицы наблюдений. Мерами дистанции могут служить: мера Минковского, мера расстояния по Евклиду, супремум-норма или расстояние Чебышева, меры сходства Жаккара и Сьеренсена, коэффициент корреляции Пирсона и многие другие меры [7].

Кластерный анализ включает набор алгоритмов типологизации, группирующих данные в наглядные структуры-таксоны: иерархическая древовидная кластеризация, двухходовое объединение, метод K -средних и другие [8]. Наиболее часто используется иерархический алгоритм "Дендрограмма", версии которого отличаются правилами вычисления расстояний между кластерами.

Критерием корректности типологизации является устойчивость результата относительно выбора алгоритма кластерного анализа. Проверяют устойчивость, применяя несколько различающихся алгоритмов. Если результаты содержательно близки, то полученная типология корректна. В ином случае необходимо предположить, что задача кластерного анализа не имеет решения и в реальности корректной типологии не существует.

Кроме иерархических методов типологизации получили распространение итерационные процедуры, с помощью которых пытаются найти наилучшее разбиение, ориентируясь на заданный критерий оптимизации, не строя при этом полного дерева (метод K -средних, алгоритмы "Форель", "Медиана", "Краб" и т.д.). В этом случае итерационный процесс начинается, как правило, со случайно выбранных кластеров, а затем путём вариации принадлежности объектов к различным кластерам решается двухаспектная задача:

- минимизации изменчивости внутри кластеров;
- максимизации изменчивости между кластерами.

В задачах анализа и прогнозирования пожарной обстановки весьма перспективно сочетание кластерного анализа с другими количественными методами (например, с корреляционным, регрессионным и факторным).

Уточняя формальную постановку задачи кластерного анализа провинций Вьетнама по пожарной обстановке, отметим, что практически она заключается в том, чтобы на основании нормализованных данных, содержащихся в двумерной матрице размером $n \times f$ (где n – число провинций, f – число социально-экономических, климатических характеристик, оперативно-служебных параметров противопожарных служб и иных показателей пожарной обстановки, разбить множество провинций n на m кластеров (достаточно однородных подгрупп) Q_1, Q_2, \dots, Q_m , так, чтобы каждый объект принадлежал одной и только одной подгруппе разбиения. При этом провинции, принадлежащие одному и тому же кластеру, должны быть сходными, в то время как провинции, принадлежащие разным кластерам, – разными.

Кластер имеет следующие *математические характеристики*: *центр, радиус, среднеквадратическое отклонение, размер кластера*:

- *центр кластера* – это среднее геометрическое место точек в пространстве переменных;

- *радиус кластера* – максимальное расстояние точек от *центра кластера*;

- *размер кластера* может быть определён либо по *радиусу кластера*, либо по *среднеквадратическому отклонению* провинций для этого кластера.

Принято следующее правило – провинция относится к кластеру, если расстояние от объекта до *центра кластера* меньше *радиуса кластера*. Если это условие выполняется для двух и более кластеров, объект является *спорным*. Очевидно, что такая неопределённость может быть устранена экспертом или аналитиком, хорошо разбирающимся в практических аспектах пожарной обстановки в стране.

Наряду с необходимостью решения задачи предварительной *стандартизации* переменных, то есть приведения значений всех характеристик к единому диапазону значений, нередко решается задача определения коэффициента важности или веса, который отражает значимость соответствующей характеристики провинции. Как правило, в качестве весов выступают экспертные оценки, полученные в ходе опроса специалистов предметной области.

Авторы оперировали равнозначными коэффициентами важности по каждой из характеристик z_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, f$) пожарной обстановки, а их стандартизация осуществлялась по формуле:

$$z_{ij}^* = \frac{z_{ij} - \bar{z}_j}{\sigma_j},$$

где \bar{z}_j – среднее значение j -й характеристики в стране

$$\bar{z}_j = \frac{1}{n} \cdot \sum_{i=1}^n z_{ij};$$

σ_j – её среднеквадратическое отклонение

$$\sigma_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2}.$$

Отбор показателей для типологизации провинций Вьетнама

Исходя из результатов анализа факторов, определяющих пожарную обстановку в регионах Вьетнама, для решения задачи их типологизации по пожарной обстановке авторы опирались на схему, отражающую факторный комплекс детерминации пожаров в стране (рис. 1).

В обоснование факторного комплекса легли результаты экспертных процедур по отбору практиками и научными работниками показателей, наиболее полно характеризующих пожарную обстановку в провинциях Вьетнама, а также результаты работ [9, 10].

Исследования авторов с использованием корреляционного анализа показали, что характеристики, детерминирующие пожарную обстановку в жилом секторе и секторе хозяйствующих субъектов, различаются, пересекаясь в некоторой своей части. При этом рассмотрена динамика характеристик с 2006 г. по 2015 г., а также их усреднённый показатель за те же годы, по которому и производилась типологизация.



Рис. 1. Факторный комплекс детерминации пожаров во Вьетнаме

Применительно к жилому сектору, при решении этой задачи методом кластеризации рассматривалась матрица размером 63 провинции, 27 характеристик (табл. 1); при решении той же задачи применительно к сектору хозяйствующих субъектов – матрица – 63 провинции, 18 характеристик (табл. 2).

Учитывая, что факторы, определяющие пожарную обстановку в названных секторах, во многом различаются, задача типологизации решалась отдельно применительно к каждому из них, затем решалась агрегированная задача.

Общий алгоритм решения задачи типологизации состоял из девяти последовательных этапов (рис. 2), каждый из которых связан с предыдущими этапами обратными связями, позволяющими на любом из них производить корректировку алгоритма. При существовании приемлемой типологизации территорий по пожарной обстановке указанные на рис. 2 этапы могут повторяться необходимое число раз (с учётом мнений экспертов) для корректировки как меры расстояний между кластерами, методов кластеризации провинций, так и системы исходных показателей и, в приемлемых границах (объяснённых с аналитической и практической точек зрения), – конечного содержания кластеров.

В процессе реализации методов кластерного анализа в современной науке применяются два основных алгоритма обработки входных данных: сравнение объектов, исходя из признаков, – *Q*-тип анализа; сравнение признаков, исходя из характеристик объектов, – *R*-тип анализа. Авторами при построении типологии территорий Вьетнама применён гибридный тип анализа (*RQ*-анализ).

На этапе отбора показателей для устранения сильно связанных из них (дублирующих друг друга), характеризующих пожарную обстановку на территориях Вьетнама, использован *R*-тип анализа. А именно, методами кластерного анализа в табл. 1 и 2 были найдены сильно связанные показатели (абсолютная величина коэффициента корреляции больше 0,8).

Таблица 1

Показатели факторного комплекса пожарной обстановки в жилом секторе

Номер показателя	Усреднённые показатели по жилому сектору (2006-2015 гг.)
1	Средний ущерб на 1 пожар, тыс. долл.
2	Частота пожаров на 1 объект жилого сектора, ‰
3	Доля электрифицированного жилья, ‰
4	Количество погибших на 1 пожар, чел.
5	Количество погибших на тыс. человек, 10³ чел.
6	Ежемесячный средний доход в госсекторе, \$
7	Доля городского населения, ‰
8	Средняя температура января, °С
9	Средняя температура июля, °С
10	Среднее количество осадков в январе, мм
11	Среднее количество осадков в июле, мм
12	Общая численность населения, млн чел
13	Доля населения в возрасте до 20 лет, ‰
14	Площадь региона, единица, млн км²
15	Площадь лесов, млн км²
16	Количество многоквартирных домов, тыс. домов
17	Количество домов частного сектора, млн домов
18	Количество пожарных-бойцов
19	Количество пожарных автомобилей
20	Количество пожаров
21	Количество погибших на пожарах
22	Количество раненых на пожарах
23	Ежегодный ущерб от пожаров, \$ млн
24	Ежегодный доход, \$ млрд
25	Ежегодный товарооборот, \$ млрд
26	Развитость автомобильных дорог, дорог/км ²
27	Площадь торговых центров, тыс. м ²

Таблица 2

Показатели факторного комплекса пожарной обстановки в секторе хозяйствующих субъектов

Номер показателя	Усреднённые показатели по жилому сектору (2006-2015 гг.)
1	Средний ущерб на 1 пожар, \$ тыс.
2	Частота пожаров на один объект надзора, ‰
3	Доля объектов первого класса пожарной опасности, ‰
4	Доля объектов второго класса пожарной опасности, ‰
5	Доля объектов третьего класса пожарной опасности, ‰
6	Ежемесячный средний доход в госсекторе, \$
7	Доля городского населения, ‰
8	Количество погибших в среднем на один пожар, чел.
9	Средняя температура января, °С
10	Средняя температура июля, °С
11	Среднее количество осадков в январе, мм
12	Среднее количество осадков в июле, мм
13	Средний размер штрафа на объект надзора I класса, \$ тыс.
14	Средний размер штрафа на объект надзора II класса, \$ тыс.
15	Средний размер штрафа на объект надзора III класса, \$ тыс.
16	Среднее количество штрафов на объект надзора I класса
17	Среднее количество штрафов на объект надзора II класса
18	Среднее количество штрафов на объект надзора III класса



Рис. 2. Блок-схема алгоритма решения задачи типологизации провинций Вьетнама по комплексным характеристикам пожарной обстановки

На рис. 3, применительно к табл. 2, показана метрика расстояний Чебышева между показателями с использованием метода медиан или взвешенного центроидного метода (weighted pair-group centroid method), позволяющего достаточно отчётливо выявить сильные связи.

Из дендрограммы, приведенной на рис. 3, следует, что показатели под номерами 16-18, а также показатели 13, 14 являются сильно связанными (в табл. 2 эти две группы выделены цветом). Оставляя из каждой группы по одному показателю, размерность пространства, характеризующего пожарную обстановку на территориях для хозяйствующих субъектов, может быть снижена с 18 до 14. Экспертами из указанных сильно связанных шести показателей были выбраны два – "Средний размер штрафа на объект надзора I класса" и "Среднее количество штрафов на объект надзора I класса". В результате типологизация в окончательном виде производилась в пространстве практически не связанных, а также слабо связанных признаков (абсолютные значения коэффициентов корреляции не выше 0,4).

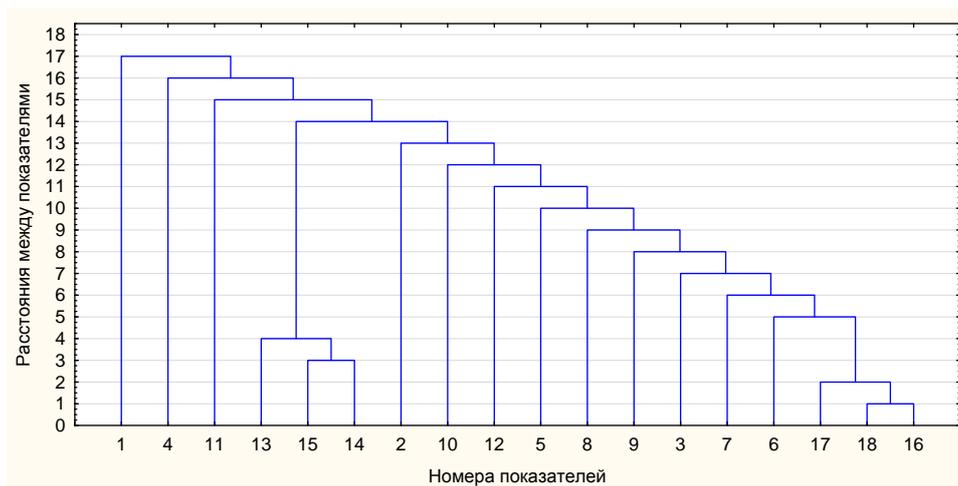


Рис. 3. Дендрограмма связей показателей пожарной обстановки в провинциях Вьетнама по сектору хозяйствующих субъектов (взвешенный центроидный метод, метрика расстояния Чебышева)

Применительно к сектору хозяйствующих субъектов при решении задачи типологизации методом кластеризации рассматривалась матрица размером 63 провинции, 14 характеристик, являющаяся частью табл. 2.

Выбор меры расстояния и метода кластеризации применительно к характеристикам пожарной обстановки осуществлялся с использованием статистического программного пакета StatSoft STATISTICA 12.5.192.7 выпуска 2014 г.

Решение задачи отбора показателей, характеризующих пожарную обстановку применительно к объектам жилого сектора, было также связано с нахождением методами кластерного анализа сильно связанных из них (величина коэффициента корреляции больше 0,8).

На рис. 4 показана метрика расстояний Чебышева между показателями с использованием взвешенного центроидного метода.

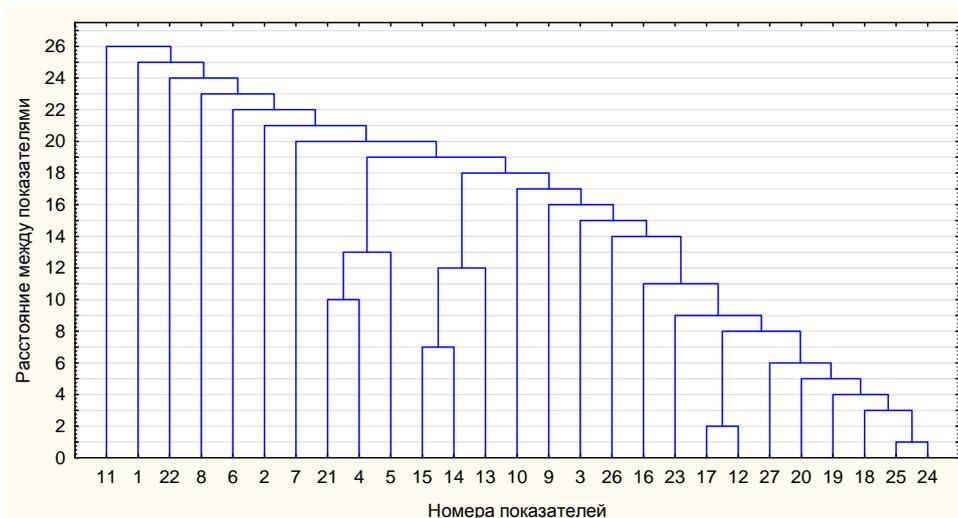


Рис. 4. Дендрограмма связей показателей пожарной обстановки в провинциях Вьетнама по жилому сектору (взвешенный центроидный метод, метрика расстояния Чебышева)

Из дендрограммы, приведенной на рис. 4, следует, что показатели под номерами 4, 5, 21; 12, 17; 13-15; 24, 25 сильно связаны (в табл. 1 выделены цветом). Это позволило снизить размерность пространства, характеризующего пожарную обстановку на территориях применительно к жилому сектору, с 27 до 21. Экспертами из указанных сильно связанных 10 показателей были выбраны 4 – "Ежегодный товарооборот", "Общая численность населения", "Количество погибших, приходящихся на один пожар", "Площадь лесов".

Таким образом, типология территорий строилась исходя из слабой связанности признаков (абсолютные значения коэффициентов корреляции не выше 0,4).

Отобранная в настоящей статье система показателей применительно к сектору хозяйствующих субъектов и жилому сектору оказалась весьма информативной, позволив содержательно и логически эффективно описать типологию территорий Вьетнама по пожарной обстановке.

Выводы

1. Чтобы учесть территориальные аспекты пожарной обстановки в регионах Вьетнама, необходимо системно подходить к рассмотрению факторов, её определяющих (социально-экономических, демографических, организационно-управленческих и материально-технических – относящиеся к деятельности противопожарной службы и других), существующих в тесном взаимодействии и взаимовлиянии.

2. Многообразие исходных признаков, характеризующих пожарную обстановку, приводит к необходимости решения задач отбора наиболее существенных из них, то есть снижения размерности пространства детерминант пожарной обстановки.

3. Среди методов многомерной классификации, предназначенных для разделения совокупности объектов (провинций Вьетнама) на однородные группы (типы) в смысле схожести условий по пожарной обстановке, важное место занимает кластерный анализ. В результате его применения возникает **наглядное представление типов территорий по пожарной обстановке**, позволяющее более точно обосновывать и принимать решения о предупреждении и эффективном тушении пожаров.

4. В процессе реализации методов кластерного анализа применяются два основных алгоритма обработки входных данных: сравнение объектов, исходя из признаков – Q -тип анализа; сравнение признаков, исходя из характеристик объектов – R -тип анализа. Наиболее эффективен при построении типологии гибридный тип анализа (RQ -анализ).

Литература

1. **Волокитина А.В., Софронов М.А.** Классификация и картографирование растительных горючих материалов. Новосибирск: изд-во СО РАН, 2002. 314 с.
2. **Рубцов А.В., Сухинин А.И., Ваганов Е.А.** Системный анализ погодной пожарной опасности при прогнозировании крупных пожаров в лесах Сибири // Исследование Земли из космоса. № 3, 2010. С. 62-70.
3. **Швецов Е.Г., Пономарев Е.И., Сухинин А.И.** Исследование пространственных и временных соотношений между молниевыми разрядами и лесными пожарами на территории Восточной Сибири // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. № 2. 2007. С. 13-18.
4. **Тростянский С.Н., Зенин Ю.Н., Минаев В.А. и др.** Оценка вероятности пожаров на основе математической модели, учитывающей факторы, определяющие долю нарушителей требований пожарной безопасности среди собственников объектов // Пожарная безопасность. № 2. 2013. С. 86-91.
5. **Loupián E.A., Mazurov A.A., Flitman E.V. and other.** Satellite Monitoring of Forest Fires in Russia at Federal and Regional Level // Mitigation and Adaptation Strategies for Global Change. №11. 2006. Pp. 113-145.
6. **Дюран Б., Оддел П.** Кластерный анализ. М.: Статистика, 1977. 128 с.
7. **Григорьев А.А.** Меры сходства в кластеризации // Известия РЭУ им. Г.В. Плеханова. Вып. 11. 2013.
8. **Гришакина Н.И., Дмитриева В.С., Манова Н.В. и др.** Многомерные статистические методы. Часть IV. Кластерный анализ: учеб.-метод. пособие. Великий Новгород: изд-во Новгородского государственного университета им. Ярослава Мудрого, 2005. 54 с.
9. **Минаев В.А., Тростянский С.Н., Чу Куок Минь.** Оценка вероятности возникновения пожаров при нарушении требований пожарной безопасности // Технологии техносферной безопасности. Вып. 5 (51). 2013. С. 15-21. <http://ipb.mos.ru/ttb/2013-5>.
10. **Минаев В.А., Фаддеев А.О.** "Медленные" катастрофы, здоровье и безопасность населения // Матер. 15-й науч.-техн. конф. "Системы безопасности –2006". М.: Академия ГПС МЧС России, 2006. С. 14-17.