



Определение классов станций



Александр КУТЫРКИН
Alexander V. KUTYRKIN

Юрий ПАЗОЙСКИЙ
Yuri O. PAZOYSKIY



Станислав ШАТСКИЙ
Stanislav O. SHATSKIY

Кутыркин Александр Васильевич — доктор технических наук, профессор Московского государственного университета путей сообщения (МИИТ).

Пазойский Юрий Ошарович — доктор технических наук, профессор, заведующий кафедрой «Железнодорожные станции и узлы» МИИТ.

Шатских Станислав Олегович — ассистент МИИТ.

Задачу классификации железнодорожных станций на основании балльной оценки объема их работы следует решать с помощью математического аппарата, в частности применяя способ, построенный на кластерном анализе. В статье представлены пошаговые алгоритмы структуризации и определения границ рассматриваемых классов.

Ключевые слова: кластерный анализ, классификация, железнодорожная станция, балльная оценка, алгоритм, структуризация, блок-схема.

Суть решаемой в данном случае проблемы структуризации заключается в разбиении исследуемого множества железнодорожных станций (представленных суммой баллов, исчисленной по показателям в условных единицах) на группы «похожих» в определенном смысле подразделений, называемых классами.

1. АЛГОРИТМ КЛАССИФИКАЦИИ

Возможность представления состояния каждого подразделения с помощью единственного показателя — суммарной балльной оценки позволяет не только сгруппировать станции в классы, но и определить границу между этими классами.

Группирование станций в классы по их схожести помогает лучше понять природу анализируемых объектов, выявить типологизационную структуру изучаемой совокупности объектов и выработать для каждого класса общие принципы управления, обслуживания и разряды оплаты труда.

Число групп (классов), на которое требуется разбить исследуемую совокупность станций, задано и равно шести: внеклассная группа, I группа (класс), II группа (класс), III группа (класс), IV группа (класс), V группа (класс).

Результатом применения данной методики должно явиться не только выделение групп станций с определением принадлежности каждой только к одному из классов, что можно осуществить с помощью классического кластерного анализа, но и вычленение границ между классами. Потребность найти границы между классами делает проблему структуризации естественным обобщением известной задачи кластеризации.

2. АЛГОРИТМ СТРУКТУРИЗАЦИИ

Пусть результаты суммарной балльной оценки станций железнодорожного транспорта представлены в виде матрицы – столбца:

$$\left. \begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{matrix} \right\} \rightarrow \left. \begin{matrix} m_1 \\ m_2 \\ \vdots \\ m_i \\ \vdots \\ m_n \end{matrix} \right\},$$

где x_i – i -ая железнодорожная станция, $i=1; n$, n – общее число станций, m_i – суммарная балльная оценка (mark) i -й станции железнодорожного транспорта.

Предлагаемый алгоритм структуризации содержит следующие основные шаги.

Шаг 1 [инициализация].

Задается общее число классов – k (в нашем случае $k=6$), на которое требуется разбить исследуемую совокупность железнодорожных станций.

Из общего множества станций $\{x_i\}$, где $i=1; n$, выбирается k станций, которые считаются начальными «центрами» классов. При этом каждый класс отождествляется с единственным начальным «центром».

Балльные оценки, соответствующие выбранным начальным «центрам», являются начальными центроидами и имеют следующие обозначения:

$$m^1[1]; m^2[1]; \dots; m^j[1]; \dots; m^k[1], \quad (1)$$

где $[\alpha]$ – α итерация, α – номер итерации, $\alpha=1; 2; 3; \dots; \alpha_{\max}$.

Задается максимальное число итераций (перестроек классов) – $\alpha_{\max}$.

Выбор начальных центроидов может содержать:

- выбор первых k -станций;
- случайный выбор k -станций.

Все исследуемые объекты распределяются по k -классам, исходя из критерия близости к одному из центроидов. Близость i -станции к j -му начальному центроиду определяется как модуль разности двух оценок

$$|m^j[1] - m_i|. \quad (2)$$

Шаг 2 [определение новых центроидов классов].

Определяются новые k – центроиды классов, которые вычисляются как средние значения оценок станций, отнесенных к сформированным классам:

$$m^j[\alpha] = \frac{1}{n_j[\alpha]} \sum_{i \in K_j[\alpha]} m_i, \quad (3)$$

где $m^j[\alpha]$ – новый центроид класса $K_j[\alpha]$, сформированного на α -итерации,

$j=1; 6$, $n_j[\alpha]$ – число объектов в классе $K_j[\alpha]$, $\alpha=2; 3; \dots; \alpha_{\max}$.

Шаг 3 [перестройка классов].

Осуществляется перераспределение станций относительно новых центроидов $m^j[\alpha]$, исходя из критерия близости к этим центроидам:

$$|m^j[\alpha] - m_i|, \quad (4)$$

где $j=1; 6$, $\alpha=2; 3; \dots; \alpha_{\max}$, $i=1; n$, n – общее число исследуемых станций.

Шаг 4 [проверка условий окончания формирования классов].

Если центроиды стабилизировались, то есть все станции относятся к классу, которому они принадлежали до текущей перестройки, или число итераций равно максимальному $\alpha_{\max}$, то переход к шагу 5, а иначе возвращение к шагу 2.

Шаг 5 [определение границ между классами].

Пусть K_j^* и K_{j+1}^* соседние классы, полученные на шаге 4 и расположенные в порядке возрастания суммарной балльной оценки, то есть для любой станции из K_j^* оценка будет меньше, чем из K_{j+1}^* , где $j=1; 5$. Выберем станцию с максимальной балльной оценкой в K_j^* и обозначим значение этой оценки через $\max_{i \in K_j^*} m_i$. Рассмотрим также станцию с минимальной балльной оценкой в K_{j+1}^* и обозначим её значение через $\min_{i \in K_{j+1}^*} m_i$. Тогда граница между классами K_j^* и K_{j+1}^* будет иметь значение





Рис. 1. Определение первичных центров тяготения.

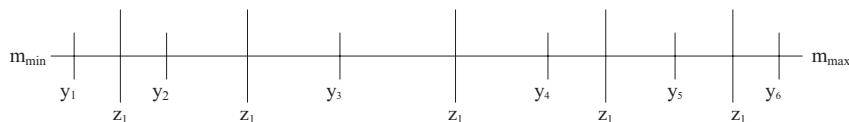


Рис. 2. Определение первичных границ кластеров.

$$\frac{1}{2} \left(\max_{i \in K_j^*} m_i + \min_{i \in K_{j+1}^*} m_i \right). \quad (5)$$

Шаг 6 [проверка качества структуризации].

При полноценной структуризации должны быть получены сильно отличающиеся средние значения оценок станций для всех классов.

3. АЛГОРИТМ ОПРЕДЕЛЕНИЯ ГРАНИЦ КЛАССОВ

Исходной информацией для определения границ классов является одномерный массив суммы баллов показателей, характеризующих работу железнодорожной станции (балльная оценка).

Для наиболее точного установления границ классов расчет должен быть выполнен для всех станций сети (генеральная совокупность станций). При упорядочивании выборки по возрастанию или убыванию алгоритм кластеризации может быть упрощен по сравнению с алгоритмом, приведенным ранее. В этом случае алгоритм структуризации имеет следующий вид.

Шаг 1. Упорядочивается выборка станций из генеральной совокупности по количеству баллов, характеризующих объемы работы.

Шаг 2. Число классов принимается равным шести: внеклассные и станции с 1 по 5 класс.

Определяются первоначальные центры тяготения кластеров.

Для этого выбираются шесть случайных чисел, упорядоченных по возрастанию — α_k (рис. 1).

$$0 < \alpha_k < 1, k = \overline{1, 6}. \quad (6)$$

Первоначальные центры тяготения кластеров находятся по формуле:

$$(m_{\max} - m_{\min}) \cdot \alpha_k = y_k, \quad (7)$$

где $m_{\min}, m_{\max}$ — соответственно минимальная и максимальная величина балла, характеризующего объем работы станций, входящих в выборку.

Шаг 3. Находятся первоначальные границы кластеров (рис. 2) по формуле:

$$\frac{y_k + y_{k+1}}{2} = z_j, \quad (8)$$

где $k = \overline{1, 6}; j = \overline{1, 5}$.

Шаг 4. Устанавливается принадлежность каждой станции к определенному кластеру:

если $m_i < z_1$, то i -я станция относится к 6-му кластеру;

если $z_1 \leq m_i < z_2$, то i -я станция соответствует 5-му;

если $z_2 \leq m_i < z_3$, то i -я станция относится к 4-му;

если $z_3 \leq m_i < z_4$, то i -я станция — к 3-му;

если $z_4 \leq m_i < z_5$, то i -я станция — к 2-му;

если $z_5 \leq m_i$, то i -я станция — к 1-му кластеру;

$i = \overline{1, n}$,
где n — число станций выборки.

Шаг 5. Определяются новые центры тяготения:

$$\frac{\sum_{m_i < z_1}^{l_6} m_i}{l_6} = y_1, \quad (9)$$

$$\frac{\sum_{z_1 \leq m_i < z_2}^{l_5} m_i}{l_5} = y_2, \quad (10)$$

$$\frac{\sum_{z_3 \leq m_i < z_3}^{l_4} m_i}{l_4} = y_3, \quad (11)$$

$$\frac{\sum_{z_3 \leq m_i < z_4}^{l_3} m_i}{l_3} = y_4, \quad (12)$$

$$\frac{\sum_{z_4 \leq m_i < z_5}^{l_2} m_i}{l_2} = y_5, \quad (13)$$

$$\frac{\sum_{m_i \geq z_5}^{l_1} m_i}{l_1} = y_6, \quad (14)$$

где l_j – число станций, попавших в i -й кластер.

Шаг 6. Проверка конца цикла.

Если полученные центры тяготения не изменились, то это говорит о сходимости алгоритма установления границ классов станций. Первому кластеру соответствует 5-й класс железнодорожных станций, второму – 4-й, третьему – 3-й класс, четвертому – 2-й класс, пятому – 1-й класс, шестому кластеру – внеклассные станции.

Если же хотя бы один из полученных центров тяготения изменил свое значение, то переходим к шагу 2 алгоритма. И так далее – до тех пор, пока не установится сходимость кластеров.

Блок-схема алгоритма определения границ классов и отнесения железнодорожных станций к классам представлена на рис. 3.

Работа алгоритма проверена на примере выборки, содержащей 782 станции

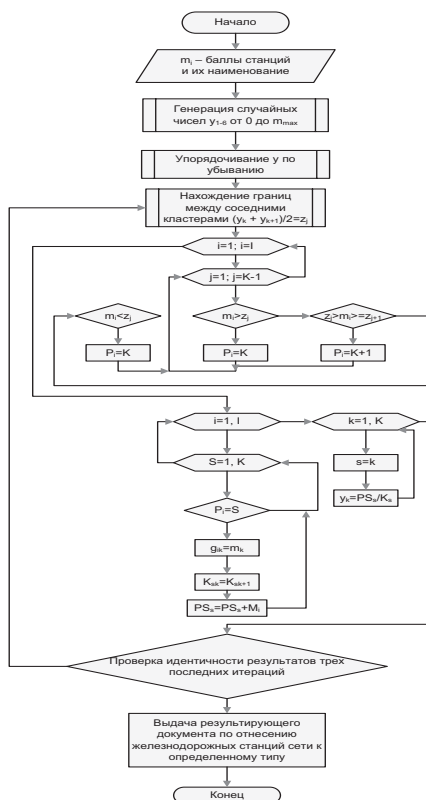


Рис. 3. Блок-схема определения границ классов и отнесения железнодорожных станций к классам.

сети железных дорог. И были определены границы классов для данной выборки и класс каждой станции. В конечном счете показана работоспособность предложенной методики и алгоритма.

ЛИТЕРАТУРА

1. Положение о железнодорожной станции. Утверждено ОАО «РЖД» 31.05.2011 г. № 1186 р.
2. Батурин А. П., Минаков А. Н., Шмулевич М. И. Организация работы полигонов железных дорог. – М.: Маршрут, 2009.

DEFINITION OF A STATION CLASS

Kutyrykin, Alexander V. – D.Sc. (Tech), professor of Moscow State University of Railway Engineering (MIIT).
Pazoykiy, Yuri O. – D.Sc. (Tech), professor, head of the department of railway stations and junctions of Moscow State University of Railway Engineering (MIIT).
Shatskyh, Stanislav O. – assistant lecturer of Moscow State University of Railway Engineering (MIIT).

The authors argue that the task of classification of the stations on the basis of scoreboard of their work content should be resolved with the help of mathematical apparatus, particularly with the technique based on cluster analysis. They give step-by-step algorithms of structuring and definition of the borders of classes under review.

Key words: cluster analysis, classification, railway station, scoreboard, rates, algorithm, structuring, flow-chart.

Координаты авторов (contact information): Кутыркин А.В. – 8(499) 241-92-83, Пазойский Ю.О. – 8(495)684-23-65, Шатских С.О. – SShatskih@gmail.com

