

Mikhail N. Pavlenkov,
ScD, (Doctor in economics), professor;

Pavel M. Voronin,
ScD, (Economics);

Tatyana V. Zhuravleva,
Postgraduate,
Institute for Economics and Entrepreneurship
NNGUn.a. NI. Lobachevskiy

The Method of Cluster Analysis Estimation of Investment Projects

Key words: Cluster, investment, tasks, algorithm.

Annotation: This article contains research of cluster analysis of investment projects, algorithms and tasks of cluster analysis.

В условиях ограниченности финансовых средств, требующихся для развития сферы твердых бытовых отходов муниципальных образований, а также отдельных целевых программ, необходимо привлекать ресурсы за счет инвестирования, создания условий для привлечения капитала, в том числе и иностранного на территории города. Привлечение инвестиций является наиболее эффективным способом экономического развития отрасли (8).

Каждое муниципальное образование имеет свои особенности, поэтому нет единого подхода к прогнозированию развития и привлечения инвестиции в развитие (4,5). Учитывая опыт и практику работы многих муниципальных образований можно систематизировать требования к инвестиционным проектам и сформулировать их в общем виде:

- первоочередные проекты;
- проекты с гарантированным финансированием (муниципальным, региональным, федеральным);
- коммерческие проекты.

Для распределения проектов по этим группам в каждом муниципальном образовании могут использоваться различные подходы для их классификации. Следует отметить, что во множестве первоочередных проектов могут быть отнесены проекты с гарантированным финансированием, а также коммерческие проекты.

Для распределения инвестиционных проектов по этим группам можно использовать кластерный анализ (1,2,3).

Достоинство кластерного анализа заключается в том, что с его помощью производить разбиение различных объектов по целому набору признаков. Кластерный

анализ не требует никаких ограничений на вид рассматриваемых объектов и позволяет рассматривать множество исходных данных практически произвольной природы (6,7).

Как и любой другой метод, кластерный анализ имеет ограничения: количество кластеров зависит от выбираемых критериев разбиения; могут теряться индивидуальные черты отдельных объектов за счет замены их характеристиками обобщенных значений параметров кластера.

В кластерном анализе считается, что:

выбранные характеристики допускают в принципе желательное разбиение на кластеры;

единицы измерения (масштаб) выбраны правильно.

Выбор масштаба играет большую роль. Как правило, данные нормализуют вычитанием среднего и делением на стандартное отклонение, так что дисперсия оказывается равной единице.

Кластерный анализ позволяет выделить и разбить множество проектов на ряд классов (групп). Он используется, когда исходные данные представлены в виде матриц близости или расстояний между объектами, либо в виде точек в многомерном пространстве.

Выбор расстояния между проектами является важным моментом классификации, так как от этого во многом зависит окончательный вариант разбиения проектов на классы.

В настоящее время разработаны различные алгоритмы кластерного анализа, один из которых называется эталонным: на множестве объектов задается несколько исходных зон, с которых начинает работу алгоритм.

После задания эталонов алгоритм производит классификацию, иногда меняя определенным способом эталоны.

Неэталонные алгоритмы работают по иному принципу. К ним можно отнести иерархические алгоритмы кластерного анализа, процедура разрезания и др.

Пусть множество $P = \{P_1, P_2, \dots, P_k\}$ обозначает K проектов, которые отобраны для инвестирования. Для распределения данного множества на кластеры, используем методы экспертных оценок. Каждый эксперт i ($i=1, 2, \dots, N$) оценивает проект P_k , т.е. распределяет проекты по классам. В рассматриваемой задаче выделяем 3 класса: класс M_1 - первоочередные проекты; класс M_2 - проект с гарантированным финансированием; класс M_3 - коммерческие проекты.

В зависимости от исследуемых объектов для их оценки могут быть использованы различные шкалы. Для оценки проектов используем классификационную шкалу. В этой шкале числа используются как метки.

Эксперт i ($i=1, 2, \dots, N$) каждому проекту ставит в соответствии:

- 1, если он считает, что проект относится к классу M_1 ;
- 2, если он считает, что проект относится к классу M_2 ;
- 3, если он считает, что проект относится к классу M_3 ;

Результаты оценки i -м экспертом проекта P_j – обозначим через x_{ij} , а вектор $X_j = [x_{ij}]$ отвечает каждому ряду измерений для j -го проекта.

Для множества P проектов имеется множество векторов $X=\{X_1, X_2, \dots, X_k\}$, которые описывают множество P .

Пусть m – целое число ($m < K$). На основании данных множества векторов X , множество проектов P разбивается на m кластеров (подмножества) $\pi_1, \pi_2, \dots, \pi_m$ так, чтобы каждый проект P_j , принадлежал одному и только одному подмножеству разбиения (кластеру).

Задачи кластерного анализа по определению числа классов, на которые требуется разбить совокупность исследуемых объектов, можно подразделить на:

- задачи, в которых заранее задано число кластеров;
- задачи, в которых требуется определить число кластеров;
- задачи, в которых число кластеров неизвестно, а требуется построить дерево исследуемой совокупности объектов.

Рассмотрим алгоритм кластерного анализа (число кластеров задано).

Шаг 1. Расчет средних величин всей совокупности данных.

Шаг 2. Расчет среднеквадратических отклонений каждого признака по совокупности данных.

Шаг 3. Расчет матрица нормированных разностей по каждому признаку.

Шаг 4. Расчет Евклидовых расстояний между каждой парой сочетаний единиц совокупности

$$R_{ij} = \sqrt{\sum_{k=1}^k (x_{ik} - x_{jk})^2}. \quad (1)$$

Шаг 5. Определение минимального значения R_{ij} .

Шаг 6. Отнесение проектов с наименьшими значениями R_{ij} в один кластер.

Шаг 7. Расчет новых усредненных значений всех признаков для объединенного кластера.

Шаг 8. Расчет новых нормированных разностей признаков между определенным кластером и остальными. Включение этих различий в матрицы вместе значений объединившихся единиц совокупности.

Шаг 9. Расчет новых Евклидовых расстояний объединенного кластера от остальных единиц или кластеров.

Шаг 10. Выбор наименьшего из Евклидовых расстояний.

Шаг 11. Объединение в кластеры прекращается при достижении заданного числа кластеров.

Шаг 12. Оценка качества классификации. После завершения классификации необходимо оценить результаты с помощью функционалов качества. Наилучшим

разбиением считается такое, при котором достигается экстремальное (min, max) значение функционала.

В качестве таких функционалов используется:

А. Сумма квадратов расстояний до центров кластеров (Φ_1)

$$\Phi_1 = \sum_{m=1}^M \sum_{j \in \pi_m} d^2(X_j, \bar{X}_m), \quad (2)$$

где m-номер кластера (m=1,2,...M);

$\bar{X}$ -центр m-го кластера;

X_j -вектор значений переменных для j-го проекта, входящего в m-й кластер;

$d(X_j, \bar{X}_m)$ – расстояние между j-м проектом и центром m-го кластера.

Наилучшим является разбиение, при котором значение Φ_1 минимально.

В. Сумма внутрикластерных расстояний между объектами (Φ_2)

$$\Phi_2 = \sum_{m=1}^M \sum_{j \in \pi_m} d_{jm}^2, \quad (3)$$

Наилучшим считается разбиение, при котором достигается минимальное значение Φ_2 .

С. Сумма внутрикластерных дисперсий (Φ_3)

$$\Phi_3 = \sum_{m=1}^M \sum_{j \in \pi_m} D_{mj}^2, \quad (4)$$

где D_{mj} – дисперсия j-й переменной в m-м кластере.

Оптимальным считается разбиение, при котором сумма внутрикластерных дисперсий будет минимальным.

Для оценки качества разбиений используются и некоторые другие простейшие приемы. Если групповые средние существенно отличаются от общего среднего значения, то это может являться признаком хорошего разбиения. Оценка существенности различий может быть выполнена с помощью t-критерия Стьюдента.

Исследование алгоритма кластерного анализа позволяет распределить инвестиционные проекты $P = \{P_1, P_2, \dots, P_K\}$ на три группы:

- первоочередные проекты ($P^1 = \{P_1, P_2, \dots, P_{K^1}\}$);
- проект с гарантированным финансированием ($P^2 = \{P_1, P_2, \dots, P_{K^2}\}$);
- коммерческие проекты ($P^3 = \{P_1, P_2, \dots, P_{K^3}\}$).

Применение разработанной методики оценки и отбора инвестиционных проектов в целях оптимального использования финансовых ресурсов, позволяет совершенствовать управление инвестиционной деятельностью сферы твердых бытовых отходов муниципального образования, реализовывать стратегические приоритеты развития данной сферы с учетом экономических, социальных и экологических императивов.

References:

1. Ayvazian SA, Bezhaeva ZI, Staroverov OV. *Classification of multidimensional observations*. Moscow, Publishing House of the "Statistics", 1974; 238.
2. Mandel ID. *Cluster analysis*. Moscow, Finance and Statistics, 1988; *Multivariate statistical analysis*: ed. VN. Timashevicha. Moscow, UNITY, 1999.
3. Maeva LS, Pavlenkov MN, Voronin PM *Some questions Estimates of municipal solid waste of the municipality: Actual problems of the economy*. Science Center "Aeterna". Ufa, Russia, 2014; 66-69.
4. Pavlenkov MN, Voronin PM, Maeva LS. *Approach to the development of a mathematical model to predict the volume of municipal solid waste of the municipality: Education and science: the current state and development prospects. Collection of scientific papers on the materials of the International scientific and practical conference in 6 parts*. Ministry of Education and Science. 2014; 118-119.
5. Pavlenkov MN. *Development of technology management solutions: Monography*. Nizhny Novgorod, Publishing House of the Volga-Vyatka Academy of Public Administration, 2005; 128.
6. Pavlenkov MN, Kulikov AL. *Methods and models of management decisions in economic systems*. Nizhny Novgorod, Publishing House VVAGS, 2004; 322.
7. *Economic Management: Environment and Ecology industrial cities: Monography*: ed. VG. Larionov. Moscow, Biblio-Globus.