

Использование кластерного анализа при формировании статистической выборки объектов стационарной торговой сети при установлении нормативов накопления твердых коммунальных отходов

*Н.В. Мензелинцева¹, Н.Ю. Карпузова², И.М. Статюха²,
О.С. Власова², Д.В. Текушин²*

¹ *Волгоградский государственный университет*

² *Волгоградский государственный технический университет*

Аннотация: Достоверность оценки количества накопленных твердых коммунальных отходов в условиях несовершенной научно-методической базы нормирования, отсутствии общепринятых методик расчета, недостаточно высокая. При экспериментальном установлении нормативов точность в значительной степени определяется объемом статистической выборки. Показана возможность использования кластерного анализа при формировании статистической выборки торговых объектов стационарной торговой сети (ТОСТС) с установлением нормативов накопления твердых коммунальных отходов. Разработана методика формирования репрезентативной статистической выборки ТОСТС при установлении нормативов накопления твердых коммунальных отходов на основе кластерного анализа.

Ключевые слова: накопление отходов, твердые коммунальные отходы, норматив образования твердых коммунальных отходов, объект стационарной торговой сети, кластерный анализ, репрезентативная выборка.

Торгово-технологический процесс торговых объектов стационарных торговых сетей (ТОСТС) сопровождается образованием и накоплением отходов, в частности, твердых коммунальных отходов (ТКО) [1-3].

К ТКО относят отходы, которые образуются в результате производственного процесса юридических лиц, индивидуальных предпринимателей, состав которых аналогичен составу отходов потребления физическими лицами, образующихся при их использовании в жилых помещениях. Нормативы накопления отходов в соответствии с современным законодательством - это удельные показатели количества отходов на единицу измерения в определенный период времени [2-4].

Однако следует отметить, что единый подход к определению нормативов накопления ТКО для различных объектов, в том числе, и ТОСТС, в настоящее время не разработан [1-3]. Нет также официальных

документов, устанавливающих виды и объем образующихся на исследуемых объектах ТКО. Регламентированные действующими нормативными документами нормативы образования отходов для объектов торговли из-за значительных изменений экономических условий, развития сферы потребления и услуг, постоянно возрастающих экологических требований при обращении с отходами, требуют переработки. Несовершенство научно-методической базы разработки нормативов для ТКО, в том числе, и доступных методик определения и расчета, приводят к невысокой достоверности полученных результатов.

При установлении нормативов накопления ТКО проводят натурные исследования, в ходе которых определяют усредненные удельные значения объемных и массовых показателей ТКО отдельно по торговой площади продовольственными и непродовольственными товарами, суточные и объемные показатели накопления ТКО по каждому виду площади. Эти показатели определяются для каждого торгового объекта рассматриваемой сети, отличающихся по целому ряду признаков, например, по общей площади, соотношению торговых площадей с продовольственными и непродовольственными товарами, перечню оказываемых услуг и т.д. Поэтому одной из наиболее важных задач является формирование достоверной статистической выборки объектов для натуральных исследований [5,6].

Установить достоверный объем выборки по одному из известных способов для ГОСТС представляется некорректным из-за специфики объектов [5,6]. В этом случае более целесообразно использовать методику кластерного анализа [6-8].

Первоначально ставится задача установления общего объема выборки. Исходя из того, что торговая сеть обычно включает в себя значительное количество объектов, логичным является использование больших выборок.

На основе экспериментов установлено, что минимальный объем выборки $N_{\text{выб}} = 30$. При формировании такой выборки следует учитывать, что торговая площадь каждого ТООТС не должна отличаться от средней площади по всей генеральной совокупности более чем на 10%.

Далее все объекты выборки делятся на типические группы на основе методов кластерного анализа.

Задача кластерного анализа состоит в делении совокупности на m кластеров на основании схожих признаков, при этом выполняются следующие условия: каждый кластер содержит идентичные объекты, а объекты, принадлежащие разным кластерам, отличаются значимо. При формировании кластеров следует учитывать, что все рассматриваемые объекты генеральной совокупности отнесены к какому-либо из кластеров, и в каждый кластер входит хотя бы один объект. Необходимо отметить, что один объект не может принадлежать двум кластерам сразу [9-11].

Если обозначить количество характерных для объекта исследований признаков как n , то, используя n -мерное пространство признаков, объект кластеризации в нем можно изобразить в виде точки. Расстояние между точками n -мерного пространства (метрика) определяет степень подобия (сходство между объектами). Чем это расстояние меньше, тем точки более схожи и наоборот.

Метрика между объектами i и k в n -мерном пространстве признаков d_{ik} должна удовлетворять условиям [6,7,12]:

1. неотрицательности, т.е. $d_{ik} \geq 0$;
2. симметрии, т.е. $d_{ik} = d_{ki}$;
3. неравенства треугольника, т.е. $d_{ik} \leq d_{ij} + d_{jk}$;
4. различимости нетождественных объектов, т.е. если $d_{ik} \neq 0$, то $i \neq k$;
5. неразличимости тождественных объектов, т.е. если $d_{ik} = 0$, то $i = k$.

В настоящее время для расчета метрики используется более 50 различных способов [6,7]. Если признаки количественные, то достаточно часто пользуются евклидовым расстоянием [6], которое можно определить по формуле (1):

$$d_{ik} = \left(\sum_{j=1}^m (x_{ij} - x_{kj})^2 \right)^{\frac{1}{2}} \quad (1)$$

где d_{ik} – расстояние между i -ым и k -ым объектами; x_{ij} и x_{kj} – численное значение j -ой переменной, соответственно для i -го и k -го объекта; m – количество переменных, характерных для объектов.

Например, для торговой сети, в которую входит N торговых объектов, сформирована генеральная совокупность Toi ($i=1,2, \dots, N$). Для любого из этих объектов определены два характерных признака: общая площадь торговых помещений (Si) и доля торговой площади продовольственными товарами (Pi). Следовательно, на двумерной плоскости каждый торговый объект сети, как объект кластеризации можно представить точкой, имеющей координаты (Si, Pi).

Генеральную совокупность делят на кластеры различными методами. Можно использовать вероятностный подход, метод k -средних, метод k -medians, EM-алгоритм и др.[13-16].

Наиболее часто применим метод k -средних, суть которого состоит в последовательной минимизации суммарного квадрата отклонений точек, принадлежащих данному кластеру, от его центров [14-16]:

$$J = \sum_{i=1}^k \sum_{x_p \in S_i} (x_p - \mu_i)^2 \quad (2)$$

где k - число кластеров, Si - полученные кластеры, $i= 1,2,\dots,k$; μ_i - центры масс векторов $x_p \in Si$.

Методом K - средних формируется K различных кластеров, которые располагаются на максимально возможных расстояниях друг от друга. Первоначально определяют K кластеров на основании выдвинутой гипотезы,

предшествующих наблюдений, результатов расчета по специальным статистическим программам. Затем, для минимизации внутрикластерной изменчивости и, наоборот, максимизации изменчивости между различными кластерами, изменяется отнесение объектов к первоначальным кластерам. Случайным образом назначаются центры новых кластеров и определяются метрики между объектами и центрами вновь сформированных кластеров. Каждый объект относят к новому наиболее близко расположенному к нему кластеру. Далее для каждого кластера определяются средние значения, принимаемые за центры тяжести новых кластеров, необходимо заметить, что их число равно числу переменных K , используемых для анализа. Описанные операции, а именно расчет метрик между каждым объектом и центрами тяжести кластеров, формирование новых кластеров и определение их центров тяжести, повторяются до момента, когда их положение в пространстве не будет меняться.

Суммируя сказанное выше, можно описать алгоритм деления на кластеры методом K -средних [7]: определение (любым описанным выше способом) K значений в качестве первичных центров кластеров; формирование промежуточных кластеров с соблюдением требования минимума расстояния от объекта до центра тяжести; нахождение кластерных средних, как центров тяжести вместо первичных; повтор последнего действия до минимизации изменений координат кластерных центров.

Допустим, что все ТОТС первоначальное разделены на три кластера ($k=3$). Следовательно, формула (2) запишется в виде:

$$J = \sum_{i=1}^3 \sum_{x_p \in S_i} (TO_i - TO_{k_i})^2 \quad (3)$$

где TO_i – центр объекта, TO_{k_i} – центры первоначальных кластеров.

Метрика (евклидово расстояние) между объектами рассчитывается:

$$d^2 (TO_i, C_k) = (S_i - S_k)^2 + (P_i - P_k)^2 \quad (4)$$

При формировании кластеров, рассматриваемый ТОСТС (TO_i) будет принадлежать кластеру C_k с минимальным расстоянием $d(TO_i, C_k)$.

Центры тяжести сформированных кластеров находят по формуле (5):

$$C_i = \frac{\sum_{j=1}^m u_{ij}x_j}{\sum_{j=1}^m u_{ij}} \quad (5)$$

где m – объекты ТОСТС, входящие в кластер C_i ; u_{ij} – коэффициент принадлежности, т.е. $u_{ij}=1$, если TO_i входит в кластер C_i , в противном случае $u_{ij}=0$.

Переформирование кластеров путем перераспределения объектов, пересчет центров тяжести осуществляют либо до момента их стабилизации, либо до момента, когда все ТОСТС останутся в своем кластере.

Затем определяют число ТОСТС N_i каждого кластера и систематизируют объекты каждого кластера $TO_j = (S_j, P_j)$ по отношению к центру тяжести окончательно сформированного кластера.

Для формирования статистической выборки находят v_i –долю ТОСТС, принадлежащих кластеру, в общем количестве ТОСТС N , входящих в кластеры, $N = N_1 + N_2 + N_3$, т.е. $v_i = N_i/N$, и определяют количество торговых объектов каждого кластера, включаемых в выборку: $n_i \approx 30v_i$. Следует подчеркнуть, что при выборе объектов рассматриваемого кластера для включения в выборку учитывается их систематизация.

Если доля ТОСТС 1-го кластера $v_1=1/5$, в статистическую выборку включают каждый 5-ый ТОСТС в порядке возрастания. Для второго кластера $v_2=1/2$, и в выборку будет отобран каждый 2-ой торговый объект опять же в порядке возрастания, и т.д. Следует помнить, что $n_1 + n_2 + n_3 = n$.

Таким образом формируется репрезентативная выборка торговых объектов стационарной торговой сети (TO_1^* , TO_2^* , ..., TO_{30}^*) для дальнейшего анализа.

Выводы

1. Показана возможность использования кластерного анализа при формировании статистической выборки торговых объектов стационарной торговой сети с установлением нормативов накопления твердых коммунальных отходов.

2. Разработана методика формирования статистической выборки объектов стационарной торговой сети при установлении нормативов накопления твердых коммунальных отходов на основе кластерного анализа.

Литература

1. Степанова И.А., Степанов А.С. Обзор систем сбора и удаления отходов в антропогенных экосистемах // Самарский научный вестник. 2020. т. 9, №2. С. 121-131.

2. Ниязгулов У.Д., Шканов С.И., Цховребов Э.С. Интеграция нормативных и организационно-технических решений в процессе обращения с ТКО и крупногабаритным мусором в населенных пунктах // Вестник ПНИПУ. Прикладная экология. Урбанистика.. 2020. №2. С. 134-148.

3. Куприн Р.Г., Цховребов Э.С., Ниязгулов У.Д. Правовое обеспечение и экономическое регулирование обращения с отходами и вторичными ресурсами // Качество. Инновации. Образование. 2018. №32(153). С. 62-70.

4. Азаров В.Н., Азаров А.В., Мензелинцева Н.В., Статюха И.М. Исследование норм накопления твердых коммунальных отходов урбанизированных территорий // Социология города. 2020. №1. С. 48-57

5. Гмурман В.Е. Теория вероятности и математическая статистика. М.: Юрайт, 2013. 173 с.

6. Городничев Р.М., Пестрякова Л.А., Ушницкая Л.А., Левина С.Н., Давыдова П.Н. Методы экологических исследований. Основы статистической обработки данных. / Под ред. Городничева Р.М. Якутск: Издательский дом СВФУ, 2019. 94 с.

7. Леончик Е.Ю. Кластерный анализ. Терминология, методы, задачи. ОНУ им. И.И. Мечникова, ИМЭМ –Изд. 2-ое, перераб. и доп. изд. Одесса: Б.в., 2011. 67 с.
 8. Мандель И.Д. Кластерный анализ. М.: Финансы и статистика, 1988. 176 с.
 9. Костенко С.А. Технология применения многомерного шкалирования и кластерного анализа // Фундаментальные исследования. 2012. №11 (часть 4) . С. 927-930.
 10. Мальсагов, А. И. Кластерный подход в региональной политике // Инженерный вестник Дона. 2012. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2012/806
 11. Романенко И. И., Романенко М.И. Стратегия развития региональной экономики через кластерные образования // Инженерный вестник Дона. 2018. № 2. URL: ivdon.ru/ru/magazine/archive/N2y2018/4818
 12. Estivill-Castro, V Why so many clustering algorithms: a position paper // ACM SIGKDD Explorations Newsletter. 2002. Vol.4(1). pp. 65-75.
 13. Tan P., Kumar V. Introduction to Data Mining. NewYork: PearsonAddison-Wesley., 2006. 125 p.
 14. Pivovarova I., Systematic Approach in Ecological Zoning// Journal of Engineering and Applied Sciences. 2015. pp. 11-15.
 15. Bandyopadhyay S., Maulik U. Anevolutionary technique based on K-means algorithm foroptimal clusteringin RN // Information Sciences. 2002. Vol.146. pp. 221–237.
 16. DaviesD.L., BouldinD.W. A cluster separation measure // IEEE Transactionson Pattern Analysis and Machine Intelligence. 1979. Vol.1(2). pp. 224-227.
-

References

1. Stepanova I.A., Stepanov A.S. Samarskiy nauchnyj vestnik. 2020. t. 9, №2. pp. 121-131.
2. Niyazgulov U.D., Shkanov S.I., Tskhovrebov E.S. Vestnik PNIPU. Prikladnaya ekologiya. Urbanistika.. 2020. №2. pp. 134-148.
3. Kuprin R.G., Tskhovrebov E.S., Niyazgulov U.D. Kachestvo. Innovatsii. Obrazovaniye. 2018. №32 (153). pp. 62-70.
4. Azarov V.N., Azarov A.V., Menzelintseva N.V., Statyukha I.M. Sotsiologiya goroda. 2020. №1. pp. 48-57.
5. Gmurman V.E. Teoriya veroyatnosti i matematicheskaya statistika [Probability theory and mathematical statistics]. M.: Yurayt, 2013. 173 p.
6. Gorodnichev R.M., Pestryakova L.A., Ushnitskaya L.A., Levina S.N, Davydova P.N. Metody ekologicheskikh issledovaniy. Osnovy statisticheskoy obrabotki dannykh [Methods of ecological research. Fundamentals of statistical data processing]. Pod red. Gorodnicheva R.M. Yakutsk: Izdatel'skiy dom SVFU, 2019. 94 p.
7. Leonchik E.YU. Klasternyj analiz. Terminologiya, metody, zadachi [Cluster analysis. Terminology, methods, tasks]. ONU im. I.I. Mechnikova, IMEM –Izd. 2-oye, pererab. i dop. izd. Odessa: B.v., 2011. 67 p.
8. Mandel' I.D. Klasternyj analiz [Cluster analysis]. M.: Finansy i statistika, 1988. 176 p.
9. Kostenko S.A. Fundamental'nyye issledovaniya. 2012. №11 (chast' 4). pp. 927-930.
10. Mal'sagov, A. I. Inzhenernyj vestnik Dona. 2012. № 2. URL: ivdon.ru/ru/magazine/archive/n2y2012/806
11. Romanenko I. I., Romanenko M.I. Inzhenernyj vestnik Dona. 2018. № 2. URL: ivdon.ru/ru/magazine/archive/N2y2018/4818



12. Estivill-Castro, V ACM SIGKDD Explorations Newsletter. 2002. Vol.4 (1). pp. 65-75.
13. Tan P., Kumar V. Introduction to Data Mining. NewYork: PearsonAddison-Wesley., 2006. 125 p.
14. Pivovarova I. Journal of Engineering and Applied Sciences. 2015. pp. 11-15.
15. Bandyopadhyay S., Maulik U. Information Sciences. 2002. Vol.146. pp. 221–237.
16. DaviesD.L., BouldinD.W. IEEE Transactionson Pattern Analysis and Machine Intelligence. 1979. Vol.1 (2). pp. 224-227.