

УДК 004.942

Мацелюх Ю.Р., аспірант кафедри інформаційних систем та мереж, спеціальність 124 Системний аналіз

(Національний університет «Львівська політехніка», м. Львів, Україна)

МЕТОДИКА ВИБОРУ МЕТОДУ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ МЕТРИК ЯКОСТІ ДЛЯ ОПТИМІЗАЦІЇ ПАСАЖИРСЬКИХ ПЕРЕВЕЗЕНЬ У РОЗУМНОМУ МІСТІ

В умовах зростання міського населення та розвитку транспортних систем, оптимізація пасажирських перевезень набуває особливої актуальності для забезпечення ефективного функціонування розумних міст. Кластеризація даних є потужним інструментом для виявлення закономірностей та структуризації інформації про пасажиропотоки, що дозволяє покращити планування маршрутів, розкладів та розподіл ресурсів у громадському транспорті. Однак, вибір оптимального методу кластеризації для конкретної задачі є нетривіальним завданням, що вимагає врахування характеристик даних та цілей аналізу.

Сучасні дослідження у галузі системного аналізу [1-3] підтверджують, що інтеграція методів кластеризації до системи управління транспортом сприяє створенню адаптивних, ефективних та низьковуглецевих моделей перевезень [4, 5]. Фундаментальні основи методів кластеризації в оптимізації пасажирських перевезень громадським транспортом сформовані відомими науковцями як в Україні, так і за її межами [6-10].

Проведені дослідження присвячені розробці методики вибору методу кластеризації на основі метрик оцінки якості, який дозволяє систематизувати процес вибору та підвищити ефективність застосування кластеризації для задач оптимізації пасажирських перевезень. Запропонована методика включає наступні етапи:

1. Визначення цілей та задач кластеризації. На цьому етапі формулюються конкретні цілі застосування кластеризації в контексті пасажирських перевезень, наприклад, зонування маршрутів, виявлення патернів пасажиропотоків, оптимізація графіків руху. Визначаються характеристики даних, такі як обсяг, розмірність, тип ознак та наявність шуму.

2. Аналіз методів кластеризації та їх властивостей. На цьому етапі розглядаються різні категорії методів кластеризації (ієрархічні, розділяючі, на основі щільності, сіткові, нейронні мережі) та їх властивості, зокрема, чутливість до шуму, форма кластерів, масштабованість, необхідність визначення кількості кластерів, інтерпретованість результатів.

3. Попередній вибір потенційних методів кластеризації. На основі цілей, задач, характеристик даних та властивостей методів, обираються 2-3 потенційно найбільш відповідні методи для подальшого аналізу.

4. Застосування обраних методів кластеризації та метрик оцінки якості. На цьому етапі для кожного обраного методу виконуються наступні кроки:

- застосування методу кластеризації до набору даних пасажирських перевезень;
- визначення оптимальної кількості кластерів k для кожного методу, використовуючи метрики оцінки якості кластеризації.

5. Оцінка якості кластеризації за допомогою метрик. Для оцінки якості кластеризації на цьому етапі пропонується використовувати наступні метрики:

- метод ліктя, відповідно до якого будується графік залежності суми квадратичних відстаней SSE від кількості кластерів k . Оптимальна кількість кластерів визначається як "лікоть" на графіку, де спостерігається сповільнення зменшення SSE;

- метод силуету, відповідно до якого розраховується середній коефіцієнт силуету для кожного значення k . Коефіцієнт силуету для окремої i -тої точки визначається як

середній коефіцієнт силуету, який розраховується як середнє значення s_i для всіх точок даних. Оптимальна кількість кластерів відповідає найвищому значенню середнього коефіцієнта силуету;

- індекс Калінського-Харабаша, який враховує співвідношення міжкластерної та внутрішньокластерної суми квадратів відстаней. Оптимальна кількість кластерів відповідає найвищому значенню індексу Калінського-Харабаша;

- інші метрики. Залежно від специфіки даних та задач, можуть бути використані інші метрики, такі як індекс Давіса-Болдіна, інформаційний критерій Байєса тощо.

6. Аналіз метрик та вибір оптимального методу та кількості кластерів. На цьому етапі проводиться аналіз графіків та значень метрик для різних методів кластеризації та кількості кластерів. Враховуються результати різних метрик, та обирається метод кластеризації та кількість кластерів, що забезпечують найкращу якість кластеризації відповідно до обраних метрик та цілей дослідження.

Розроблена методика дозволяє систематизувати процес вибору методу кластеризації та оптимальної кількості кластерів на основі метрик оцінки якості. Застосування даної методики для задач оптимізації пасажирських перевезень дозволить підвищити ефективність кластеризації даних та отримати більш якісні результати для підтримки прийняття рішень у сфері управління громадським транспортом розумних міст.

Список використаних джерел

1. Jain A. K., Murty M. N., Flynn P. J. Data clustering: a review. *ACM Computing Surveys (CSUR)*. 1999. Vol. 31. №3. P. 264-323. DOI:10.1145/331499.331504
2. Kaufman L., Rousseeuw P. J. *Finding groups in data: an introduction to cluster analysis*. Hoboken: John Wiley & Sons, 2009. 368 p.
3. Calinski T., Harabasz J. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*. 1974. Vol. 3. №1. P. 1-27. DOI: 10.1080/03610927408827101
4. Podlesna L., Bublyk M., Grybyk I., Matseliukh Y., Burov Y., Kravets P., Lozynska O., Karpov I., Peleshchak I., Peleshchak R. Optimization model of the buses number on the route based on queueing theory in a Smart City. *CEUR Workshop Proceedings*. 2020. Vol. 2631. P. 502–515. URL: <https://ceur-ws.org/Vol-2631/paper37.pdf> (дата звернення: 01.03.2025).
5. Koshtura D., Bublyk M., Matseliukh Y., Dosyn D., Chyrun L., Lozynska O., Karpov I., Peleshchak I., Maslak M., Sachenko O. Analysis of the demand for bicycle use in a smart city based on machine learning. *CEUR workshop proceedings*. 2020. Vol. 2631, P. 172–183. URL: <https://ceur-ws.org/Vol-2631/paper13.pdf> (дата звернення: 01.03.2025).
6. Bublyk M., Udovychenko T., Medvid R. Concept of smart specialization in the context of the development of Ukraine's economy. *Economics. Ecology. Socium*. 2019. Vol. 3. No 2. P. 55–61. DOI: 10.31520/2616-7107/2019.3.2-6
7. Kowalska-Styczeń A., Bublyk M., Lytvyn V. Green innovative economy remodeling based on economic complexity. *Journal of Open Innovation: Technology, Market, and Complexity*. 2023. V.9 (3), 100091. DOI: 10.1016/j.joitmc.2023.100091
8. Chukhray N. Shakhovska N., Mrykhina O., Bublyk M., Lisovska L. Consumer aspects in assessing the suitability of technologies for the transfer. *Computer sciences and information technologies, CSIT 2019: Proc. of Int. Sci. Tech. Conf.* 2019. V. 3. P. 142-147. DOI: 10.1109/STC-CSIT.2019.8929879
9. Шипуліна Ю.С. *Механізм управління потенціалом інноваційного розвитку промислових підприємств*. Монографія. Суми: ТОВ «Друкарський дім «Папірус»», 2021. 460 с.
10. Bublyk M.I., Dulyaba N.I., Petryshyn N.Y., Drymalovska K.V. Analysis of the development of educational and scientific activity of institutions of higher education. *Economic analysis*. 2018. №28 (1). P. 30-39.