

Віталій Кіндрат, Остап Думанський*

¹ НЛТУ України, Львів, Україна, kindrat.v@nltu.lviv.ua

² НЛТУ України, Львів, Україна, к.ф.-м.н., доцент, dumanskiy@nltu.edu.ua

Розроблення веб-платформи для систематизації та ефективного управління корпоративними даними в приватних компаніях

***Анотація.** У статті розглянуто математичну модель для просторового аналізу даних бізнес-активності, побудовану на основі алгоритму кластеризації DBSCAN та формули Гаверсина. Запропоновано підхід для виявлення областей із високою щільністю об'єктів та можливістю аналізувати ступінь насиченості певних районів. Проведено моделювання показало, що використання DBSCAN дає змогу створювати стійкі кластери навіть за нерівномірного розподілу точок.*

***Ключові слова** – кластеризація, відстань Гаверсина, просторовий аналіз даних, геоінформаційні системи, щільність даних, аналітика бізнес-активності, математичне моделювання, класифікація точок.*

Однією з актуальних проблем сучасної аналітики просторових даних є визначення закономірностей розподілу об'єктів у географічному просторі. У сфері бізнес-менеджменту така задача має прикладне значення — від аналізу концентрації точок продажу до оцінки рівня насиченості ринку в різних регіонах. Автоматизація процесу виявлення кластерів бізнес-активності дозволяє підвищити ефективність прийняття управлінських рішень, зокрема у питаннях оптимального розміщення нових об'єктів чи визначення перспективних локацій. Тому розроблення математичної моделі, що дозволяє точно визначати щільні зони розташування об'єктів на основі їхніх координат, є важливим завданням у галузі просторового аналізу даних.

Основою моделі є алгоритм DBSCAN (Density-Based Spatial Clustering of Applications with Noise), який дозволяє формувати кластери довільної форми на основі щільності точок та одночасно відокремлювати шумові дані [1]. Для визначення належності точки до певного кластера використовується поняття ϵ -

околиці — множини всіх точок, що знаходяться на відстані не більшій за ε від поточної точки. Формально це можна подати як:

$$N\varepsilon(p) = \{q \in D \mid d(p,q) \leq \varepsilon\}. \quad (1)$$

Оскільки модель працює з географічними координатами, для обчислення реальної відстані між точками застосовується формула Гаверсина, що враховує кривизну земної поверхні [2]. Вона визначає дугову відстань між двома точками на сфері за їх широтою та довготою:

$$d(p_i, p_j) = 2R \arcsin \sqrt{\sin^2(\Delta\varphi/2) + \cos\varphi_i \cos\varphi_j \sin^2(\Delta\gamma/2)}, \quad (2)$$

де R - радіус Землі, φ і γ - широта й довгота відповідних точок.

Додатково в моделі вводиться показник щільності локального середовища, який можна подати як відношення кількості точок у ε -околиці до площі цієї області:

$$\rho(p) = |N\varepsilon(p)|/\pi\varepsilon^2 \quad (3)$$

Точки з високою щільністю утворюють ядра кластерів, тоді як елементи з низькою щільністю розглядаються як шумові. Таким чином, алгоритм природно виділяє області концентрації об'єктів без впливу зовнішніх параметрів масштабу або форми. Оптимальний вибір параметрів ε і *MinPts* здійснюється емпірично або за допомогою графічного аналізу відстаней (*k-distance plot*), що дозволяє адаптувати модель до конкретного набору геоданих. Комбінація DBSCAN та метрики Гаверсина забезпечує коректне виявлення щільних груп точок у геопросторі незалежно від масштабу та форми розподілу даних. Модель ефективно відокремлює реальні центри бізнес-активності від випадкових або ізольованих точок, що дає можливість отримати об'єктивну картину розміщення об'єктів у просторі. Отримані результати можуть бути застосовані для аналізу економічної щільності, прогнозування розвитку бізнес-мереж та побудови інтелектуальних систем підтримки прийняття рішень на основі просторових даних.

Для перевірки працездатності розробленої моделі було проведено експериментальне моделювання просторового розподілу бізнес-об'єктів. У рамках симуляції створено три області підвищеної щільності, що умовно відповідають регіонам великих міст. Додатково до вибірки було внесено шумові точки, рів-

номірно розподілені по території дослідження, які імітують випадкові або не пов'язані об'єкти. На основі цих даних було застосовано алгоритм DBSCAN із параметрами $\epsilon=2.0$ км та $\text{MinPts}=5$, що дозволило ідентифікувати три стійкі кластери та відокремити неструктурований шум. Для обчислення відстані між географічними координатами використано метрику Гаверсина, яка забезпечила точність оцінки просторових зв'язків у сферичній системі координат. Отримана кластерна карта демонструє компактні групи точок навколо центрів бізнес-активності та незначну кількість розріджених об'єктів поза межами кластерів.

Результати показали, що модель здатна виявити концентрацію об'єктів, навіть у випадку нерівномірного розподілу даних. Алгоритм DBSCAN продемонстрував стійкість до шуму та здатність формувати кластери довільної форми без потреби попереднього визначення їхньої кількості. Такий підхід дозволяє шукати не лише вузли бізнес-активності, а й аналізувати рівень концентрації об'єктів у різних областях.

Висновок. У роботі було розроблено та досліджено математичну модель просторової кластеризації даних бізнес-активності на основі алгоритму DBSCAN із використанням метрики Гаверсина. Проведене моделювання показало, що запропонований підхід ефективно визначає області підвищеної щільності об'єктів і дозволяє автоматично відокремлювати шумові точки без потреби у попередньому заданні кількості кластерів. Застосування географічної метрики забезпечило коректність оцінки відстаней між точками, що особливо важливо для аналізу реальних геолокаційних даних. Отримані результати демонструють, що комбінація DBSCAN і формули Гаверсина забезпечує стабільність кластерів навіть при нерівномірному розподілі даних, а також дозволяє виявляти центри ділової активності у просторі. Розроблена модель може бути використана як основа для створення інтелектуальних аналітичних систем, здатних підтримувати процеси прийняття рішень у сфері бізнес-планування та урбаністичного аналізу. Подальші дослідження можуть бути спрямовані на адаптацію параметрів ϵ та MinPts за допомогою методів машинного навчання, що підви-

щити точність і гнучкість моделі у роботі з великими наборами просторових даних.

Список використаних літературних джерел

- [1] Ester, M., Kriegel, H.-P., Sander, J., Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise // Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). — Portland, Oregon, 1996. — P. 226–231.
- [2] Sinnott, R.W. Virtues of the Haversine // Sky and Telescope. — 1984. — Vol. 68, No. 2. — P. 159.

Development of a web platform for systematization and efficient managing of corporate data in private companies

Vitalii Kindrat, Ostap Dumanskyi

The article examines a mathematical model for spatial analysis of business activity data based on the DBSCAN clustering algorithm and the Haversine formula. The proposed approach allows identifying areas with a high density of objects and analyzing the degree of saturation in specific regions. The conducted modeling demonstrated that the use of DBSCAN enables the formation of stable clusters even under uneven point distribution.