

Л.О. Кіріченко, Д.С. Хацько, П.П. Зінченко

КЛАСТЕРИЗАЦІЯ ТРАЄКТОРІЙ БРОУНІВСЬКОГО РУХУ ЗА ДОПОМОГОЮ МАШИННОГО НАВЧАННЯ

Анотація. Стаття присвячена виявленню пасток, які захоплюють броунівську частинку, за допомогою машинного навчання. Проведено моделювання траєкторії броунівської частинки за допомогою моделі броунівського руху із дрейфом, що охоплює як вільну дифузію, так і рух у пастці. Для кластеризації часових даних використовуються метод кластеризації на основі щільності DBSCAN. Універсальність цього метода дозволяє ідентифікувати кластери без попереднього знання їхньої кількості або форми, що робить його придатним для виявлення пасток. Проведене дослідження показує, що застосування методу DBSCAN дозволяє досягнути середньої точності 95.0%.

Ключові слова: Броунівський рух з дрейфом, рух частинок у пастках, кластеризація за допомогою машинного навчання, DBSCAN

Вступ і мета

Броунівський рух виступає математичною моделлю для багатьох фізичних, хімічних і біологічних явищ [1-3]. Рух молекул у клітинах, процеси дифузії та взаємодії між біологічними компонентами є фундаментальними процесами, які визначають біологічні функції та реакції організмів. Вивчення молекулярної динаміки в клітинах за допомогою броунівського руху відкриває нові шляхи для розуміння клітинних процесів і механізмів, які впливають на функціональність живих організмів.

У контексті траєкторії броунівського руху пастками називають просторові області, в яких броунівська частинка утримується протягом певного часу. Ці пастки можуть виникати природним чином або можуть бути штучно створені в лабораторіях для детального вивчення властивостей частинок [4-6]. Дослідження динаміки броунівського руху в системах з пастками сприяє глибшому розумінню цього руху та виявленню складних закономірностей, що лежать в його основі. Виявлення пасток є окремим і складним завданням, яке було предметом різних досліджень із застосуванням як ансамблевих підходів [6,7], так і аналізу траєкторій окремих частинок [8-11]. Багато з цих методів використову-

ють ковзаючі вікна для виявлення пасток [10-12], хоча деякі підходи передбачають цілісне дослідження траєкторій частинок [8].

Для вирішення проблеми виявлення пасток було застосовані методи машинного навчання, як показано в [12], проте такі дослідження все ще перебувають на стадії розвитку. У нашій роботі ми пропонуємо застосування методу кластеризації часових рядів на основі методів машинного навчання [13-17]. Тому основною метою цього дослідження є кластеризація часової траєкторії броунівської частинки для виявлення пасток – областей, де рух частинки знає значного сповільнення.

Для виконання завдання кластеризації було обрано метод DBSCAN [17-20]. Цей метод спирається на розподіл щільності об'єктів даних у просторі, що дозволяє ідентифікувати області з високою щільністю, які можуть відповідати кластерам. Крім того, метод не залежить від апріорних знань про кількість і форму кластерів, що дозволяє ідентифікувати пастки без попередньої інформації про їхню кількість і часові інтервали.

Моделювання траєкторій частинок

Для моделювання траєкторії руху частинки було обрано модель броунівського руху з дрейфом. Модель броунівського руху з дрейфом відрізняється від стандартного броунівського руху тим, що вводить поняття дрейфу або середнього спрямованого зміщення частинки. У стандартному броунівському русі частинка випадковим чином рухається в різних напрямках без будь-якого переважного напрямку. Однак у моделі дрейфу на частинку також впливають зовнішні сили або градієнти, які надають певного напрямку її руху. Модель броунівського руху з дрейфом широко використовується для моделювання реальних процесів у різних галузях. Ці процеси охоплюють різні галузі, такі як фізика, хімія, біологія, фінанси та інженерія. Універсальність моделі броунівського руху з дрейфом робить її безцінним інструментом для розуміння складності динамічних систем та явищ [21, 22]. Наведемо означення як для стандартного броунівського руху, так і для броунівського руху з дрейфом.

Стохастичний процес $\{W(t), t \geq 0\}$ називається стандартним броунівським рухом, якщо він задовольняє наступним умовам: $W(0) = 0$; $W(t) - W(s)$ відповідає нормальному розподілу із середнім значенням 0 та дисперсією $t - s$; $W(t) - W(s)$ не залежить від $W(v) - W(u)$, якщо ці часові інтервали не перетинаються.

Броунівський рух з дрейфом $\{B(t), t \geq 0\}$ розширює поняття стандартного броунівського руху, додаючи до нього дрейфовий член. Цей дрейф μ вносить систематичний зсув або спрямовану складову в рух частинки на додаток до випадкових флуктуацій. Процес $\{B(t), t \geq 0\}$ задовольняє наступним умовам: $B(0) = 0$; має незалежні та стаціонарні прирости; процес нормально розподілений $B(t) \sim N(\mu t, t)$.

Процес броунівського руху з дрейфом, що включає обидва дрейфи μ і коефіцієнт дифузії σ , в загальному вигляді можна виразити так:

$$B_{(\mu, \sigma)}(t) = \mu t + \sigma W(t). \quad (1)$$

Для моделювання траєкторії броунівської частинки в середовищі з пастками інтервали руху частинки в пастках вибираються випадковим чином, з урахуванням конкретних обмежень, що диктуються конкретною задачею, яка розглядається. Рух частинки в пастці також моделювався за допомогою броунівського руху з дрейфом, але з істотним зменшенням швидкості частинки, що досягалося за рахунок зменшення коефіцієнта дифузії в M раз:

$$B_{(\mu, \sigma)}(t) = \mu t + \frac{\sigma}{M} W(t). \quad (2)$$

Таким чином, для моделювання траєкторії частинки використовували (1) для опису руху частинки під час вільної дифузії, а (2) – для опису її руху в межах пастки. Такий підхід дозволив отримати повне уявлення про поведінку частинки, враховуючи як необмежений броунівський рух, так і вплив пастки.

Застосування методу кластеризації DBSCAN для виявлення пасток

Використання методу кластеризації на основі щільності показало багатообіцяючий потенціал у сфері кластеризації часових рядів. У нашому дослідженні ми використали можливості методу DBSCAN (Density-Based Spatial Clustering of Applications with Noise) для ефективною кластеризації часових даних.

У нашому дослідженні застосування DBSCAN дозволило виявити складні часові патерни і структури всередині траєкторій, що поглибило наше розуміння основної динаміки системи.

Метод DBSCAN визначає кластери як області з високою щільністю, відокремлені від областей з низькою щільністю. DBSCAN здатний виявляти кластери різної форми та розміру, що робить його особливо ефективним, коли кластери мають різну щільність або складну форму.

DBSCAN використовує два основні параметри: радіус і мінімальну кількість сусідів. Починаючи з випадкової точки з набору даних, метод розглядає

всі точки в заданому радіусі від цієї точки. Якщо кількість точок у цій області перевищує мінімальну кількість сусідів, область вважається щільною і розглядається як один кластер. Потім цей процес поширюється на сусідні точки, які також задовольняють умові щільності, і так далі. Точки, які не належать до жодного кластера і не відповідають критерію щільності, вважаються шумом [13, 23].

У контексті представленого дослідження всі точки траєкторії були класифіковані як такі, що належать до одного з двох кластерів. Перший кластер охоплював точки, динаміка яких відповідала броунівському руху, тоді як другий кластер складався з точок, що демонстрували рух у пастці.

Пов'язуючи певні точки траєкторії з кластером пастки, можна було розшифрувати динаміку і тривалість утримання частинок у різних ділянках експериментальної установки. Ця часова інформація дозволила отримати повне уявлення про явище пастки та його зміни в часі.

Експеримент і результати

Дослідження фокусується на двовимірному броунівському русі, проте кластеризація проводилася незалежно для кожної координати, щоб підвищити точність виявлення пасток. Часова траєкторія моделювалася згідно з рівняннями (1) і (2). Кількість пасток, їхня тривалість та часове розташування визначалися випадковим чином, з використанням рівномірного розподілу, з дотриманням попередньо визначених обмежень. Наприклад, були накладені обмеження, щоб запобігти перекриттю позицій пасток у часовому просторі.

На рис. 1 вгорі показано ілюстрацію змодельованого двовимірного броунівського руху у фазовому просторі X та Y . Області, що відповідають пасткам, для наочності виділено червоним кольором. У нижній частині рис. 1 подано проєкції цього руху на осі X та Y , що дає повне зображення траєкторії частинки в обох координатах.



Рисунок 1 - Двовимірний броунівський рух у фазовому просторі (вгорі) та проєкції на осі X і Y (внизу)

Вхідними даними для кластеризації були модельні траєкторії з пастками, а вихідними – анотовані відрізки цієї траєкторії, пов'язані з пастками. Подальший кластерний аналіз мав на меті розбити цю траєкторію на окремі кластери, які відповідали різній динаміці, що її демонструвала частинка.

У процесі моделювання певні значення параметрів моделі використовувалися як базова лінія: $\mu = 0.1$, $\sigma = 20$ та $M = 10$. Крім того, було проведено серію експериментів, в яких значення цих параметрів систематично змінювалися.

Оскільки ми використовували реалізацію на основі моделей, ми змогли оцінити точність процесу кластеризації. Для оцінки точності роботи методів були використані матриці помилок для бінарної класифікації. На рис. 2 представлено одну з типових матриць помилок для проведеного експерименту. В даному випадку точність склала 96%. Порівнюючи передбачені кластерні розподіли з істинною природою точок траєкторії (класифіковані як такі, що належать до пастки чи ні), ці матриці дозволили нам кількісно оцінити успіх алгоритму кластеризації в правильній ідентифікації точок траєкторії, що належать і не належать до пастки.

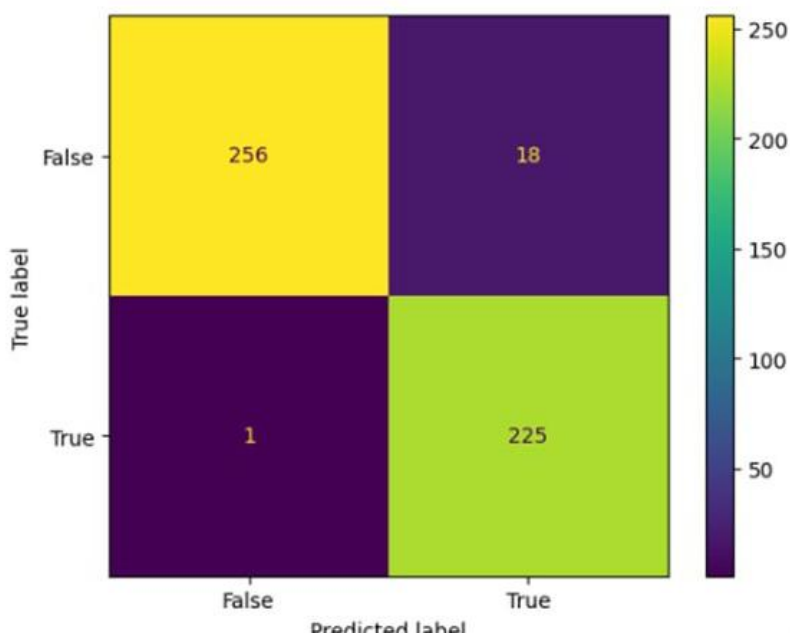


Рисунок 2 - Матриця помилок одного з результатів кластеризації

Для побудови моделі броунівського руху та її кластеризації була використана мова програмування Python, яка відома своєю великою колекцією бібліотек, пристосованих для математичних задач, машинного навчання та візуалізації результатів. Універсальність Python у наукових обчисленнях зробила його ідеальним вибором для нашого дослідження, дозволивши нам ефективно реалізувати та проаналізувати складну динаміку броунівських траєкторій руху.

На рис. 3 представлено ілюстрацію одного з остаточних результатів кластеризації, отриманих за допомогою методу DBSCAN. Виділені області зображують окремі кластери, що представляють ідентифіковані пастки в даних траєкторії.



Рисунок 3 - Результати кластеризації методом DBSCAN для проекції на ось X (вгорі) та Y (внизу)

Варто зазначити, що метод DBSCAN продемонстрував неабияку точність у визначенні точок, які потрапляли в пастки на траєкторії. Тим не менш, він був схильний до випадкової помилкової класифікації, особливо в сценаріях, де різниця між вільною дифузією і рухом, пов'язаним з пастками, була менш чіткою. Зокрема, представлений метод є схильними до помилкової класифікації сегментів вільного руху як руху, пов'язаного з пасткою. Ця відмінність є дуже

важливою, оскільки вона відображає чутливість методу до складних просторових патернів і шуму в даних.

Для оцінки ефективності методу кластеризації була проведена комплексна оцінка за допомогою серії з 100 випробувань. Аналіз виявив помітні відмінності в показниках точності. Метод DBSCAN показав середню точність 95.0%.

Моделювання та кластеризація проводилися в діапазоні значень дрейфу μ , коефіцієнта дифузії σ та коефіцієнта m . Оцінка не виявила помітної залежності ефективності методів від дрейфу μ . Як показали експерименти, точність методів знижується лише при малих значеннях σ . При менших значеннях коефіцієнта m точність кластеризації зменшується. Це явище пояснюється тим, що для малих ділянок рух частинки в пастці наближається до її безперешкодного руху. Отже, відмінні риси утримання стають малопомітними, що створює проблему для методів кластеризації ефективно виявляти такі аномалії в русі частинок.

Висновки

У цій статті досліджено моделювання траєкторій частинок на основі броунівського руху з дрейфом і вивчено ефективність методу кластеризації DBSCAN на основі машинного навчання, з метою виявлення пасток. Модель броунівського руху охоплювала як вільну дифузію, так і рух у пастках. Використання методу DBSCAN продемонструвало його потенціал для точної ідентифікації областей пасток у даних траєкторії.

Дослідження показало, що DBSCAN продемонстрував загалом високу середню точність, зокрема, досягнув середньої точності 95.0%. Комплексні експерименти, проведені в рамках цього дослідження, дозволили зрозуміти вплив параметрів моделі на результати кластеризації.

Майбутні дослідження в цьому напрямку можуть заглибитися в застосування більш досконалих методів машинного навчання для виявлення пасток і дослідити інтеграцію методів багатовимірної кластеризації для фіксації складної поведінки частинок.

ЛІТЕРАТУРА

1. E. Frey and K. Kroy, "Brownian Motion: a Paradigm of Soft Matter and Biological Physics," *Ann. d. Phys. (Leipzig)*, vol. 14, pp. 20-50, 2005. [Online]. Available: arXiv:cond-mat/0502602.

2. M.A. Pinsky and S. Karlin, Brownian Motion and Related Processes[Online]. Available:https://faculty.ksu.edu.sa/sites/default/files/an_introd_to_stoch_modeling_4th_ed.pdf.
3. L. Kirichenko and R. Lavrynenko, "Probabilistic Machine Learning Methods for Fractional Brownian Motion Time Series Forecasting," *Fractal and Fractional*, vol. 7, no. 7, pp. 517, 2023. DOI:10.3390/fractalfract7070517.
4. C. Bustamante, J. Liphardt, and F. Ritort, "The nonequilibrium thermodynamics of small systems," *Physics Today*, 2005. [Online]. Available: <https://arxiv.org/ftp/cond-mat/papers/0511/0511629.pdf>.
5. L. P. Faucheux et al., "Optical thermal ratchets," *Physical Review Letters*, vol. 74, no. 8, pp. 1504-1507, 1995. DOI: 10.1103/PhysRevLett.74.1504.
6. V. Briane et al., "A computational approach for detecting micro-domains and confinement domains in cells: A simulation study," *Phys. Biol.*, vol. 17, p. 025002, 2020. DOI 10.1088/1478-3975/ab5e1d.
7. V. Briane, V. Kervrann, and M. Vimond, "Statistical analysis of particle trajectories in living cells," *Phys. Rev. E*, vol. 97, p. 062121, 2018. DOI:10.1103/PhysRevE.97.062121.
8. Y. Lanoiselée et al., "Detecting Transient Trapping from a Single Trajectory: A Structural Approach," *Entropy*, vol. 23, p. 1044, 2021. DOI:0.3390/e23081044.
9. P. Koo and J. K. Mochrie, "Systems-level approach to uncovering diffusive states and their transitions from single-particle trajectories," *Phys. Rev. E*, vol. 94, p. 052412, 2016. DOI:10.1103/PhysRevE.94.052412.
10. N. Meilhac et al., "Detection of confinement and jumps in single-molecule membrane trajectories," *Phys. Rev. E*, vol. 73, p. 011915, 2006. DOI:10.1103/PhysRevE.73.011915.
11. A. R. Vega et al., "Multistep Track Segmentation and Motion Classification for Transient Mobility Analysis," *Biophys. J.*, vol. 114, pp. 1018–1025, 2018. DOI: 10.1016/j.bpj.2018.01.012.
12. P. Dosset et al., "Automatic detection of diffusion modes within biological membranes using back-propagation neural network," *BMC Bioinform.*, vol. 17, p. 197, 2016. DOI:10.1186/s12859-016-1064-z.
13. S. Aghabozorgi, A. Shirkhorshidi, and T. J. Wah, "Time-series clustering. A Decade Review," *Information systems*, vol. 53, pp. 16-38, 2015.
14. L. Kirichenko, O. Pichugina, and H. Zinchenko, "Clustering time series of complex dynamics by features," *CEUR Workshop Proceedings*, vol. 3132, pp. 83-93, 2022.

15. S. Khlamov et al., "Statistical modeling for the near-zero motion detection of objects in series of images from the data stream," in 12th International Conference on Advanced Computer Information Technologies (ACIT 2022), 2022, pp. 1-4. DOI: 10.1109/ACIT54803.2022.9913151.
16. L. Kirichenko, T. Radivilova, and A. Tkachenko, "Comparative analysis of noisy time series clustering," CEUR Workshop Proceedings, vol. 2362, 2019.
17. A. Alqahtani et al., "Deep Time-Series Clustering: A Review," Electronics, vol. 10, no. 23, p. 3001, 2021. DOI:10.3390/electronics10233001.
18. R. J. G. B. Campello et al., "Density-Based Clustering Based on Hierarchical Density Estimates," in Advances in Knowledge Discovery and Data Mining, PAKDD 2013, 2013, pp. 305-316. DOI:10.1007/978-3-642-37456-2_14.
19. L. Kirichenko et al., "Forecasting weakly correlated time series in tasks of electronic commerce," in Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2017, 2017, vol. 1, pp. 309-312, doi: 10.1109/STC-CSIT.2017.8098793.
20. L. McInnes and J. Healy, "Accelerated Hierarchical Density Based Clustering," in 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, USA, 2017, pp. 33-42, doi: 10.1109/ICDMW.2017.12.
21. S. Kadloor, R. S. Adve, and A. W. Eckford, "Molecular communication using Brownian motion with drift," IEEE Transactions on NanoBioscience, vol. 11, no. 2, pp. 89-99, 2012. DOI:10.1109/tnb.2012.2190546.
22. S. Zhu and C. Ravishankar, "A scalable approach to approximating aggregate queries over intermittent streams," in Proceedings. 16th International Conference on Scientific and Statistical Database Management, 2004., Santorini, Greece, 2004, pp. 85-94, doi: 10.1109/SSDM.2004.1311196.
23. S. Louhichi, M. Gzara and H. Ben Abdallah, "A density based algorithm for discovering clusters with varied density," 2014 World Congress on Computer Applications and Information Systems (WCCAIS), Hammamet, Tunisia, 2014, pp. 1-6, doi: 10.1109/WCCAIS.

Received 17.11.2023.

Accepted 21.11.2023.

Clustering Brownian motion trajectories using machine learning

The article is dedicated to detecting traps encountered by a Brownian particle based on machine learning methods. The trajectory of the Brownian particle was modeled using a drift-extended Brownian motion model, encompassing both free diffusion and particle movement within a trap. The density-based spatial clustering of applica-

tions with noise (DBSCAN) method was employed for clustering the motion trajectory. The versatility of this method allows the identification of clusters without prior knowledge of their quantity or shape, making it suitable for trap detection. The conducted research demonstrates that the application of the DBSCAN method achieves an average accuracy of 95.0%

Кіріченко Людмила Олегівна – д.т.н., професор кафедри прикладної математики Харківського національного університету радіоелектроніки.

Хацько Дарина Сергіївна – магістрантка кафедри прикладної математики Харківського національного університету радіоелектроніки.

Зінченко Петро Петрович – аспірант кафедри прикладної математики Харківського національного університету радіоелектроніки.

Kirichenko Lyudmila Olehivna - Ph.D., professor of the Department of Applied Mathematics of the Kharkiv National University of Radio Electronics.

Khatsko Daryna Serhiyivna - is a master's student at the Department of Applied Mathematics of the Kharkiv National University of Radio Electronics.

Zinchenko Petro Petrovych - is a graduate student of the Department of Applied Mathematics of the Kharkiv National University of Radio Electronics.