Символічна регресія в моделюванні та дизайні матеріалів .   
Концепція, що широко використовується у процесі виявлення нових матеріалів, відома як матеріальна станція. Вона передбачає дослідження численних комбінацій хімічних елементів, які можна визначити за даними періодичної таблиці для створення нових матеріалів, принаймні в теоретичному аспекті. Однак, окрім базової сумісності елементів, необхідно враховувати ще один ключовий критерій – стабільність матеріалу, що є визначальним фактором для оцінки можливості його існування. Символічна регресія (SR) довела свою ефективність у цій сфері [1], ставши потужним інструментом для ідентифікації складних математичних залежностей, відомих як дескриптори [2], а також для побудови нових моделей, що відображають кореляцію зі стабільністю матеріалів [3].

З іншого боку, процеси виявлення та дизайну матеріалів спрямовані на максимальне використання прогнозування їх властивостей і створення матеріалів із заданими характеристиками. Це завдання може виявитися вкрай складним, оскільки певні цільові властивості можуть проявлятися лише в специфічних кристалічних структурах. Ба більше, для досягнення високої продуктивності необхідно, щоб певні характеристики матеріалу були точно узгоджені між собою [4]. Тому критично важливо визначити параметри, що впливають на функціональність матеріалу, а також дослідити взаємозв’язки між ними, щоб оптимізувати кінцевий результат [5].

1. Xiong J, Zhang T, Shi S (2020) Machine learning of mechanical

properties of steels. Sci China Technol Sci 63(7):1247–1255.

https:// doi. org/ 10. 1007/ s11431- 020- 1599-5

2. Seko A, Togo A, Tanaka, I.. (2018). In: Tanaka I (ed) Descriptors

for machine learning of materials data. Springer, Singapore, pp

3–23. https:// doi. org/ 10. 1007/ 978- 981- 10- 7617-6\_1

3. He M, Zhang L (2021) Machine learning and symbolic regres-

sion investigation on stability of mxene materials. Comput Mater

Sci 196:110578. https:// doi. org/ 10. 1016/j. comma tsci. 2021.

110578

4. Hautier G (2019) Finding the needle in the haystack: materi-

als discovery and design through computational ab initio high-

throughput screening. Comput Mater Sci 163:108–116. https://

doi. org/ 10. 1016/j. comma tsci. 2019. 02. 040

5. Wang Y, Wagner N, Rondinelli JM (2019) Symbolic regression

in materials science. MRS Commun 9(3):793–805. https:// doi.

org/ 10. 1557/ mrc. 2019. 85

6. Weng B, Song Z, Zhu R, Yan Q, Sun Q, Grice CG, Yan Y, Yin

W-J (2020) Simple descriptor derived from symbolic regression

accelerating the discovery of new perovskite catalysts. Nat Com-

mun 11(1):1–8. https:// doi. org/ 10. 1038/ s41467- 020- 17263-9

Символічна регресія в будівництві

Починаючи з галузі цивільного будівництва, в гідравліці, де рідина протікає через труби, рівняння Колбрука для опису тертя потоку є широко відомою моделлю серед інженерів, яку також приймають і в суміжних інженерних дисциплінах [7]. У одному з досліджень було створено кілька моделей рівняння Колбрука за допомогою SR [7], причому отримані результати є досить точними. Як зазначають автори, їх підходи діють лише в умовах турбулентного режиму, оскільки перехід від ламінарного до турбулентного режиму не може бути ефективно описаний цим рівнянням. Вони доповнили свої попередні дослідження підходами, розробленими за допомогою генетичних алгоритмів та нейронних мереж, що дозволяють краще моделювати цей перехід. У подальшому дослідженні було виявлено більш прості рівняння, які поєднують як ламінарні, так і турбулентні гідравлічні режими, що зменшує потребу в окремих моделях для кожного з них [8]. Однак їх набір даних був згенерований через вибірку з уже наявних рівнянь, а не на основі експериментальних даних, що обмежує застосовність отриманих моделей [8].

З іншого боку, бетон, основний матеріал для будівництва великих споруд, також став популярною темою для досліджень з використанням AI. Багато робіт зосереджено на створенні моделей для оцінки коефіцієнта сейсмічного пікового дрейфу [9, 10], глибини проникнення в бетонні блоки [11], зсувної ємності сталевих волоконних бетонних балок, моніторингу реакції пожежі конкретних структур [12], а також сейсмічної реакції через аналіз крихкості [13]. Інші дослідження фокусуються на точному описі залишкового життя бетонних конструкцій при втомі [14, 15] або на опорі зсуву в підшипниках. Важливо зазначити, що, хоча у цих випадках моделювання зазвичай здійснюється на основі вимірювань, в деяких випадках [11, 14, 16] створені рівняння використовують звичайні, часто вживані формули.

7. Praks P, Brkić D (2018) Symbolic regression-based genetic

approximations of the colebrook equation for flow friction.

Water. https:// doi. org/ 10. 3390/ w1009 1175

8. Milošević M, Brkić D, Praks P, Litričin D, Stajić Z (2022)

Hydraulic losses in systems of conduits with flow from laminar

to fully turbulent: a new symbolic regression formulation. Axi-

oms. https:// doi. org/ 10. 3390/ axiom s1105 0198

9. Hamidia M, Ganjizadeh A (2022) Post-earthquake damage evalu-

ation of non-ductile rc moment frames using surface crack pat-

terns. Struct Control Health Monitor. https:// doi. org/ 10. 1002/ stc.

3024

10. Mansourdehghan S, Dolatshahi KM, Asjodi AH (2022) Data-

driven damage assessment of reinforced concrete shear walls

using visual features of damage. J Build Eng 53:104509. https://

doi. org/ 10. 1016/j. jobe. 2022. 104509

11. Imran Latif QBA, Memon ZA, Mahmood Z, Qureshi MU, Milad

A (2022) A machine learning model for the prediction of con-

crete penetration by the ogive nose rigid projectile. Appl Sci.

https:// doi. org/ 10. 3390/ app12 042040

12. Naser MZ (2019) Heuristic machine cognition to predict fire-

induced spalling and fire resistance of concrete structures. Autom

Constr 106:102916. https:// doi. org/ 10. 1016/j. autcon. 2019.

102916

13. Rezaei H, Zarfam P, Golafshani EM, Amiri GG (2022) Seismic

fragility analysis of rc box-girder bridges based on symbolic

regression method. Structures 38:306–322. https:// doi. org/ 10.

1016/j. istruc. 2021. 12. 058

14. Gan L, Wu H, Zhong Z (2022) Integration of symbolic regres-

sion and domain knowledge for interpretable modeling of

remaining fatigue life under multistep loading. Int J Fatigue

15:106889. https:// doi. org/ 10. 1016/j. ijfat igue. 2022. 106889

16. Ren J, Zhang L, Zhao H, Zhao Z, Wang S (2022) Determi-

nation of the fatigue equation for the cement-stabilized cold

recycled mixtures with road construction waste materials based

on data-driven. Int J Fatigue 158:106765. https:// doi. org/ 10.

1016/j. ijfat igue. 2022. 106765

Моделювання в фізиці та астрономії

Дані астрономічних спостережень є надзвичайно багатими, і методи штучного інтелекту (AI) добре розроблені для їх ефективної обробки та аналізу. Наприклад, галактичні кластери визнано одними з найбільших структур у Всесвіті [16], оскільки вони складаються з кількох галактик, що включають темну речовину, чорні дірки та інші елементи [17]. Більше того, ці структури грають важливу роль в еволюції та формуванні об'єктів, механізми яких досі не є повністю зрозумілими [17]. Враховуючи це, було розроблено різноманітні підходи для їх дослідження, такі як пошук математичних виразів для уніфікації властивостей галактичних кластерів з їх масою [16], вивчення взаємодії галактик [18], моделювання зміщення складу [19] та оцінка загальної маси субхалів [17].

Окрім цього, SR успішно застосовується для моделювання транзитної спектроскопії екзопланет [20], оскільки спостереження за планетарними транзитами в різних діапазонах хвиль є важливим методом для отримання інформації про структуру та склад атмосфери екзопланет [20]. Серед інших значущих застосувань варто зазначити відновлення орбітальних аномалій за допомогою спостережень положення та швидкості, створення моделі динаміки [21], прогнозування гравітаційних хвильових сурогатів [22], реконструкцію параметра подвійності у підходах до прогнозування космічної дистанції через сильно лінзовані гравітаційні хвильові події [23], а також аналіз сонячної активності під час сонячного циклу, що дозволяє успішно розкривати основні закони щодо управління магнітними хвилями [24].

17. Wadekar, D., Thiele, L., Villaescusa-Navarro, F., Hill, J.C.,

Cranmer, M., Spergel, D.N., Battaglia, N., Anglés-Alcázar,

D., Hernquist, L., Ho, S.: Augmenting astrophysical scaling

relations with machine learning : application to reducing the

SZ flux-mass scatter. arXiv (2022). https:// doi. org/ 10. 48550/

ARXIV. 2201. 01305

18. Shao H, Villaescusa-Navarro F, Genel S, Spergel DN, Anglés-

Alcázar D, Hernquist L, Davé R, Narayanan D, Contardo G,

Vogelsberger M (2022) Finding universal relations in subhalo

properties with artificial intelligence. Astrophys J 927(1):85.

https:// doi. org/ 10. 3847/ 1538- 4357/ ac4d30

19. Delgado AM, Wadekar D, Hadzhiyska B, Bose S, Hernquist L,

Ho S (2022) Modelling the galaxy-halo connection with machine

learning. Mon Not R Astron Soc 515(2):2733–2746. https:// doi.

org/ 10. 1093/ mnras/ stac1 951

20. Matchev KT, Matcheva K, Roman A (2022) Analytical modeling

of exoplanet transit spectroscopy with dimensional analysis and

symbolic regression. Astrophys J 930(1):33. https:// doi. org/ 10.

3847/ 1538- 4357/ ac610c

22. Barsotti D, Cerino F, Tiglio M, Villanueva A (2022) Gravita-

tional wave surrogates through automated machine learning.

Class Quantum Grav 39(8):085011. https:// doi. org/ 10. 1088/

1361- 6382/ ac5ba1

21. Manzi, M., Vasile, M.: Orbital anomaly reconstruction using

deep symbolic regression. In: 71st International Astronautical

Congress (2020)

23. Arjona R, Lin H-N, Nesseris S, Tang L (2021) Machine learning

forecasts of the cosmic distance duality relation with strongly

lensed gravitational wave events. Phys Rev D 103:103513.

https:// doi. org/ 10. 1103/ PhysR evD. 103. 103513

24. Shepherd SJ, Zharkov SI, Zharkova VV (2014) Prediction of

solar activity from solar background magnetic field variations

in cycles 21–23. Astrophys J 795(1):46. https:// doi. org/ 10. 1088/

0004- 637x/ 795/1/ 46

Символічна регресія в медицині

Більшість застосувань символічної регресії (СР) в медичній науці слідують подібній структурі, починаючи з виявлення ключових характеристик в даних (зазвичай вимірюваннях або анамнезі пацієнтів) і завершуючи створенням моделей для прогнозування, прогнозів або діагностики. Наприклад, у дослідженні, що стосується хвороби Паркінсона [25], СР змогла виявити важливі фактори, що стосуються змін у ході пацієнтів. Подібним чином, в іншому дослідженні [26] була побудована зрозуміла модель ризику для прогнозування виживаності хворих на рак молочної залози. Наступні застосування включають аналіз вимірювань для діагностики гепатоцелюлярної карциноми (раку печінки) [27], оцінку рівня гемоглобіну та глюкози в крові, моделювання ключових особливостей із пальців [28], а також поєднання пацієнтів з подібними характеристиками для покращення променевої терапії через реконструкцію дози та оптимізацію процедур [29]. Інші приклади застосувань включають аналіз вимірювань з різних частин тіла для моделювання ходьби людини [30].

Крім того, використання СР допомогло покращити попередні моделі, надавши математичну інтерпретацію до традиційних методів машинного навчання, які зазвичай мають статус «чорної коробки» [31]. Зокрема, СР була використана для створення математичних формул для перетворення коваріатів з медичних записів пацієнтів, що в подальшому дозволило застосувати ці формули в моделі Кокса [32]. Це призвело до підвищення точності прогнозування в порівнянні з використанням лише моделі Кокса. В іншому дослідженні, присвяченому вагітності та розвитку прееклампсії, СР показала кращі результати, порівняно з моделями, заснованими на логістичній регресії, виявляючи важливі взаємозв'язки між ключовими ознаками [33].

25. Hughes, J.A., Houghten, S., Brown, J.A.: Gait model analysis of

parkinson’s disease patients under cognitive load. In: 2020 IEEE

Congress on Evolutionary Computation (CEC), pp. 1–8 (2020).

https:// doi. org/ 10. 1109/ CEC48 606. 2020. 91856 21

26. Alaa AM, Gurdasani D, Harris AL, Rashbass J, van der Schaar

M (2021) Machine learning to guide the use of adjuvant therapies

for breast cancer. Nat Mach Intell 3(8):716–726. https:// doi. org/

10. 1038/ s42256- 021- 00353-8

27. Goyal, R.: A symbolic regression approach to hepatocellular car-

cinoma diagnosis using hypermethylated cpg islands in circulat-

ing cell-free dna. medRxiv (2022). https:// doi. org/ 10. 1101/ 2022.

01. 25. 22269 799

28. Golap MA-U, Raju SMTU, Haque MR, Hashem MMA (2021)

Hemoglobin and glucose level estimation from ppg characteris-

tics features of fingertip video using mggp-based model. Biomed

Signal Process Control 67:102478. https:// doi. org/ 10. 1016/j.

bspc. 2021. 102478

29. Virgolin, M., Alderliesten, T., Bel, A., Witteveen, C., Bosman,

P.A.N.: Symbolic regression and feature construction with gp-

gomea applied to radiotherapy dose reconstruction of childhood

cancer survivors. In: Proceedings of the Genetic and Evolution-

ary Computation Conference, pp. 1395–1402. Association for

Computing Machinery, (2018). https:// doi. org/ 10. 1145/ 32054 55.

32056 04

30. Dasgupta P, Hughes JA, Daley M, Sejdić E (2021) Is human

walking a network medicine problem? An analysis using sym-

bolic regression models with genetic programming. Comput

Methods Progr Biomed 206:106104. https:// doi. org/ 10. 1016/j.

cmpb. 2021. 106104

31. Wilstrup C, Cave C (2022) Combining symbolic regression with

the cox proportional hazards model improves prediction of heart

failure deaths. BMC Med Inform Decis Mak 22(1):1–7. https://

doi. org/ 10. 1186/ s12911- 022- 01943-1

32. Cox DR (1972) Regression models and life-tables. J R Statistical

Soc: Ser B (Methodol) 34(2):187–202. https:// doi. org/ 10. 1111/j.

2517- 6161. 1972. tb008 99.x

33. Wilstrup, C., Hedley, P.L., Rode, L., Placing, S., Wøjdemann,

K.R., Shalmi, A.-C., Sundberg, K., Christiansen, M.: Sym-

bolic regression analysis of interactions between first trimes-

ter maternal serum adipokines in pregnancies which developpre-eclampsia.

medRxiv (2022). https:// doi. org/ 10. 1101/ 2022.

06. 29. 22277 072