



Аппроксимация обменно- корреляционных функционалов в квантовой химии математическими выражениями с помощью символьной регрессии

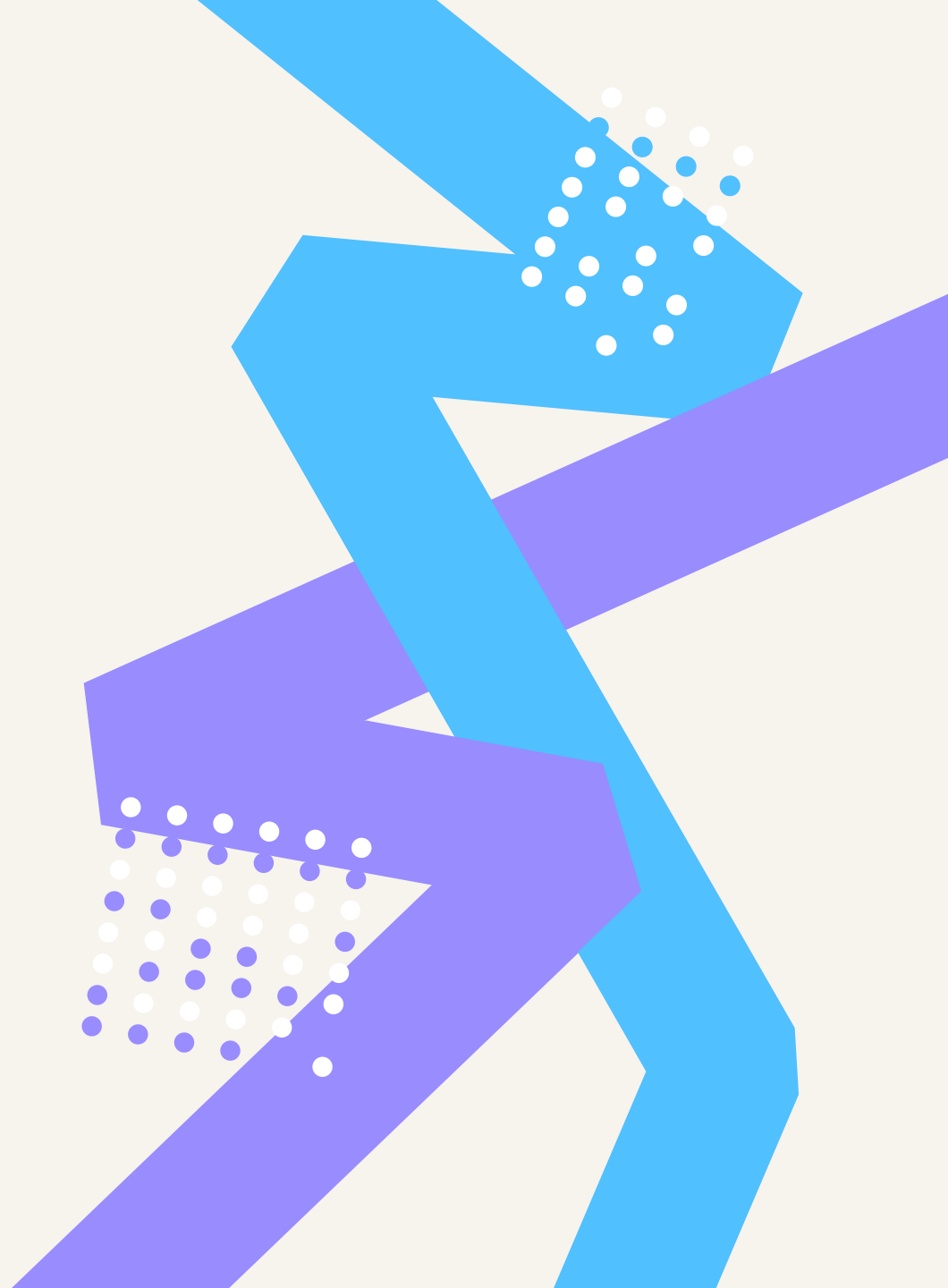
Выполнили:

Ким Павел

Кулаев Кирилл

Куратор:

Рябов Александр



Содержание

- 01 Введение
- 02 Обзор методов
- 03 Экспериментальная часть
- 04 Выводы

01



Введение



Актуальность

$$\hat{H}\Psi = \left[\hat{T} + \hat{V} + \hat{U} \right] \Psi = \left[\sum_{i=1}^N \left(-\frac{\hbar^2}{2m_i} \nabla_i^2 \right) + \sum_{i=1}^N V(\mathbf{r}_i) + \sum_{i<j}^N U(\mathbf{r}_i, \mathbf{r}_j) \right] \Psi = E\Psi$$

Стационарное уравнение Шредингера

$$E[\rho] = T_s[\rho] + \int d\mathbf{r} v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) + E_H[\rho] + E_{\text{xc}}[\rho]$$

Полная энергия Кона-Шэма

$$v_{\text{eff}}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + e^2 \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' - \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})}$$

the corresponding energy expression, are the only unknowns in the Kohn–Sham approach to density functional theory

Потенциал Кона-Шэма

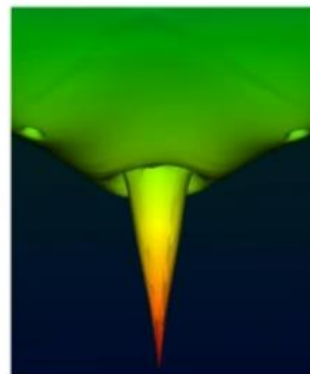
$\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N_e})$
Many-body wavefunction

Configuration interaction

$\rho(\mathbf{r}) \xleftrightarrow{v_{\text{xc}}[\rho(\mathbf{r})]} v_{\text{xc}}(\mathbf{r})$

Inverse DFT

- ❖ PDE-constrained optimization
- ❖ Complete finite-element basis
- ❖ Cusp correction
- ❖ Correct asymptotics

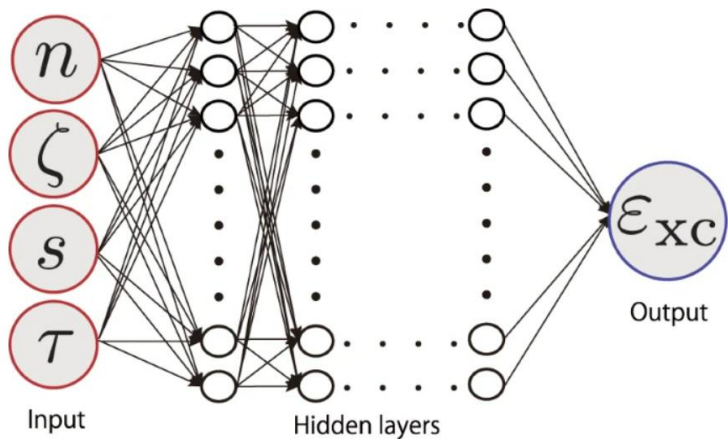


LDA functionals

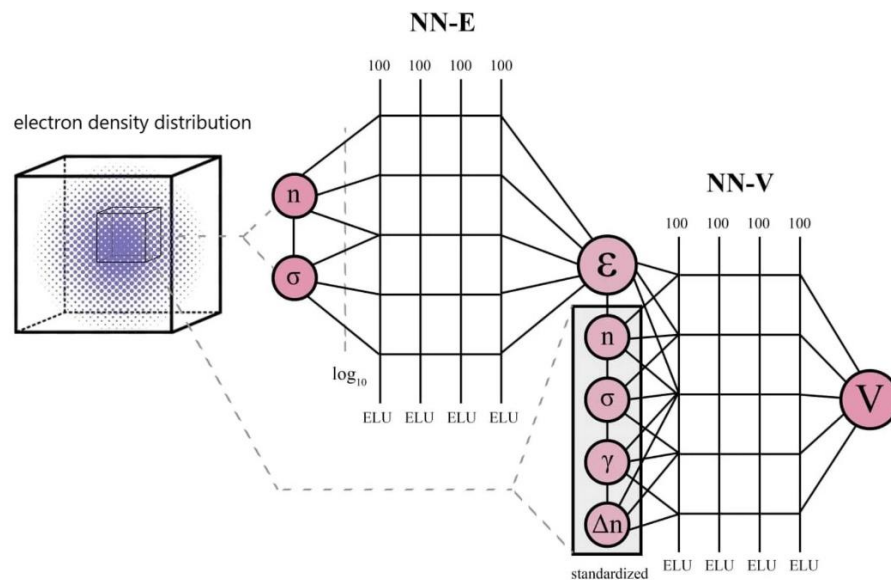
LDA exchange

- LDA_X (id=1): Slater exchange
 - P. A. M. Dirac., Math. Proc. Cambridge Philos. Soc. 26, 376 (1930) (doi: 10.1017/S0305004100016108)
 - F. Bloch., Z. Phys. 57, 545 (1929) (doi: 10.1007/BF01340281)
- LDA_X_1D_EXPONENTIAL (id=600): Exchange in 1D for an exponentially screened interaction
 - N. Helbig, J. I. Fuks, M. Casula, M. J. Verstraete, M. A. L. Marques, I. V. Tokatly, and A. Rubio., Phys. Rev. A 83, 032503 (2011) (doi: 10.1103/PhysRevA.83.032503)
- LDA_X_1D_SOFT (id=21): Exchange in 1D for a soft-Coulomb interaction
 - N. Helbig, J. I. Fuks, M. Casula, M. J. Verstraete, M. A. L. Marques, I. V. Tokatly, and A. Rubio., Phys. Rev. A 83, 032503 (2011) (doi: 10.1103/PhysRevA.83.032503)
- LDA_X_2D (id=19): Slater exchange
 - P. A. M. Dirac., Math. Proc. Cambridge Philos. Soc. 26, 376 (1930) (doi: 10.1017/S0305004100016108)
 - F. Bloch., Z. Phys. 57, 545 (1929) (doi: 10.1007/BF01340281)
- LDA_X_ERF (id=546): Short-range LDA exchange with error function kernel (erfc)
 - P. M. W. Gill, R. D. Adamson, and J. A. Pople., Mol. Phys. 88, 1005-1009 (1996) (doi: 10.1080/00268979609484488)
 - J. Toulouse, A. Savin, and H.-J. Flad., Int. J. Quantum Chem. 100, 1047-1056 (2004) (doi: 10.1002/qua.20259)
 - Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, and K. Hirao., J. Chem. Phys. 120, 8425-8433 (2004) (doi: 10.1063/1.1688752)
- LDA_X_RAE (id=549): Rae self-energy corrected exchange
 - A.I.M. Rae., Chem. Phys. Lett. 18, 574 - 577 (1973) (doi: 10.1016/0009-2614(73)80469-5)

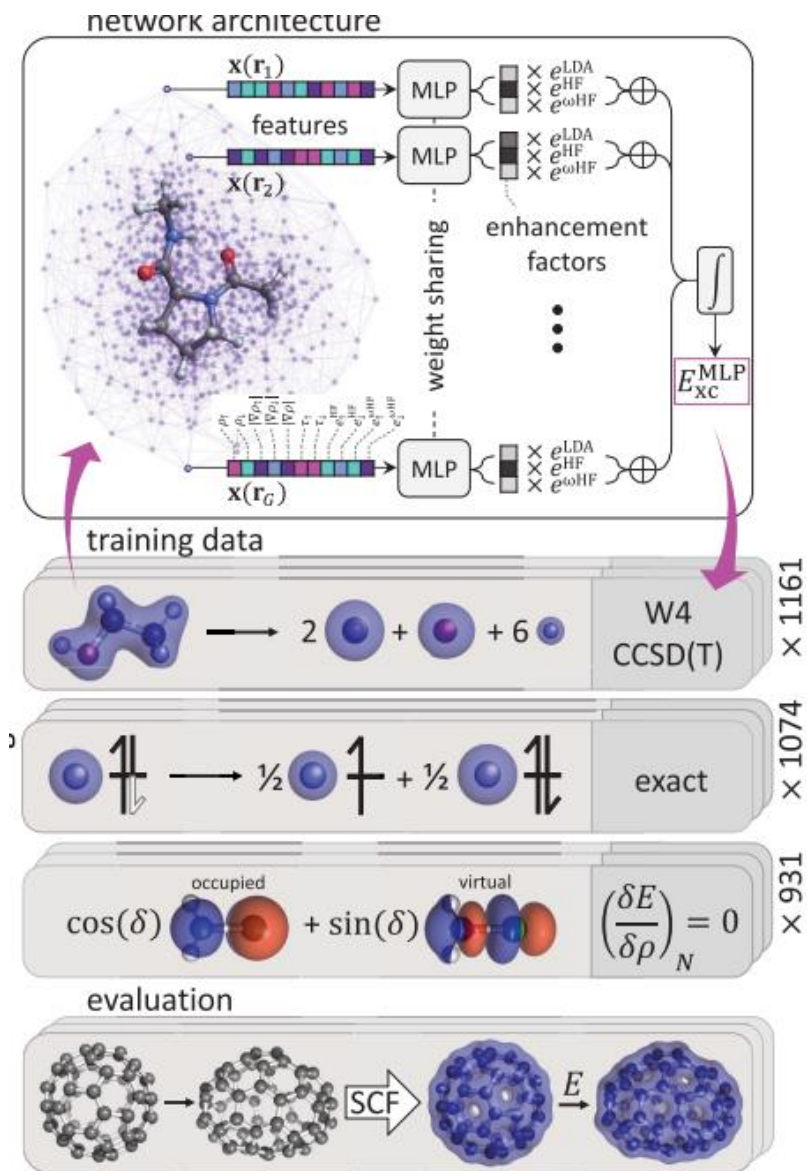
Мотивация



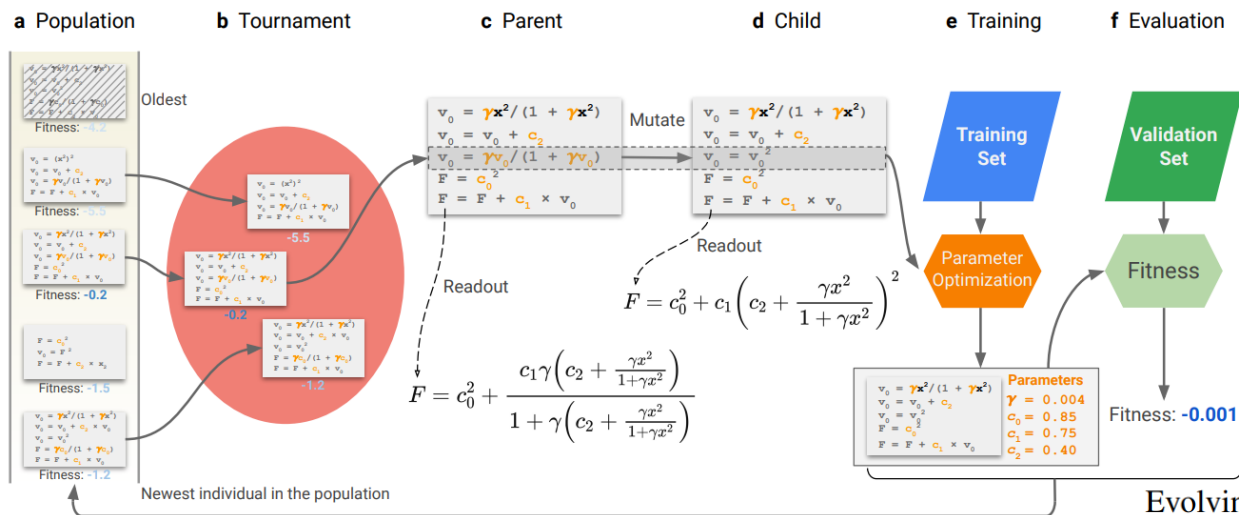
<https://doi.org/10.1038/s41524-020-0310-0>



<https://doi.org/10.1038/s41598-022-18083-1>



[DOI: 10.1126/science.abj6511](https://doi.org/10.1126/science.abj6511)



Evolving symbolic density functionals

<https://doi.org/10.48550/arXiv.2203.02540>

He Ma,¹ Arunachalam Narayanaswamy,¹ Patrick Riley,¹ Li Li,^{1*}

¹Google Research, Mountain View, CA 94043, USA

*To whom correspondence should be addressed; E-mail: leeley.lili@gmail.com

Постановка задачи

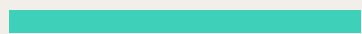
→ Цель:

Найти аналитический вид существующего нейросетевого обменно-корреляционного функционала, используя модели символьной регрессии

→ Задачи:

- Получить аналитические выражения известных обменно-корреляционных функционалов с помощью существующих алгоритмов символьной регрессии и оценить их эффективность;
- Аппроксимировать существующий нейросетевой обменно-корреляционный функционал с помощью моделей символьной регрессии, оценить качество полученного выражения;
- Сделать расчет полной энергии модельной системы в основном состоянии (молекула H_2O), используя полученный аналитический вид обменно-корреляционного потенциала, и сравнить со значением, рассчитанным *ab initio*.

02



Обзор методов



Генетический алгоритм gplearn

- + Есть возможность задавать произвольные базисные функции
- + Возможность контролировать сложность аппроксимирующей формулы
- ✗ Получение лучшего в сравнении с имеющимися решения, а не оптимального

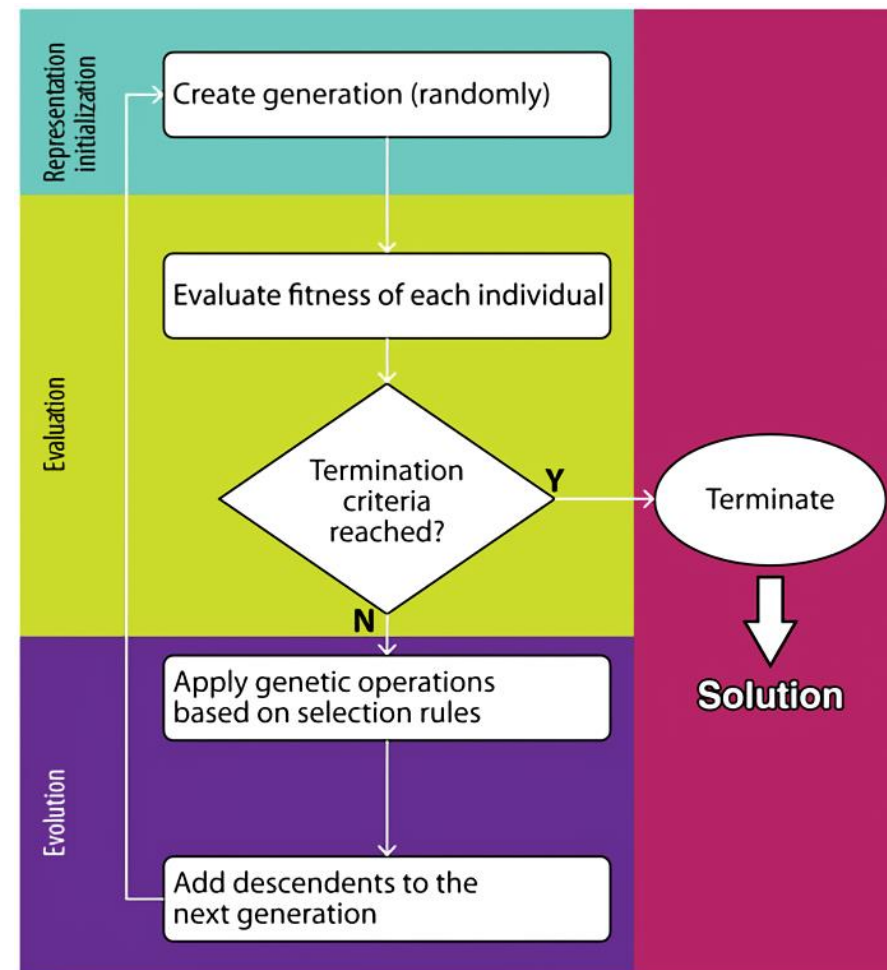
Использовали библиотеку gplearn



Genetic Programming in Python,
with a scikit-learn inspired API:

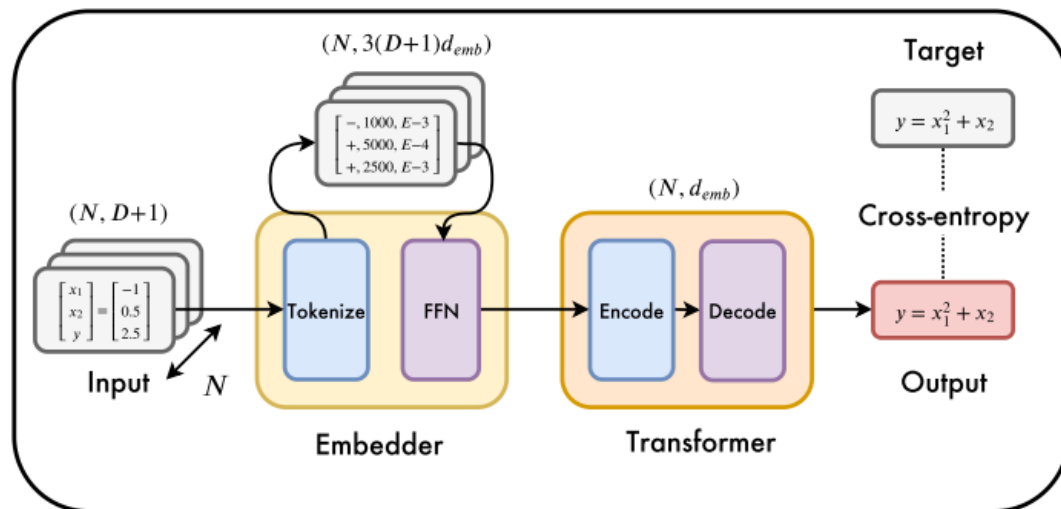
gplearn

<https://gplearn.readthedocs.io/en/stable/index.html>

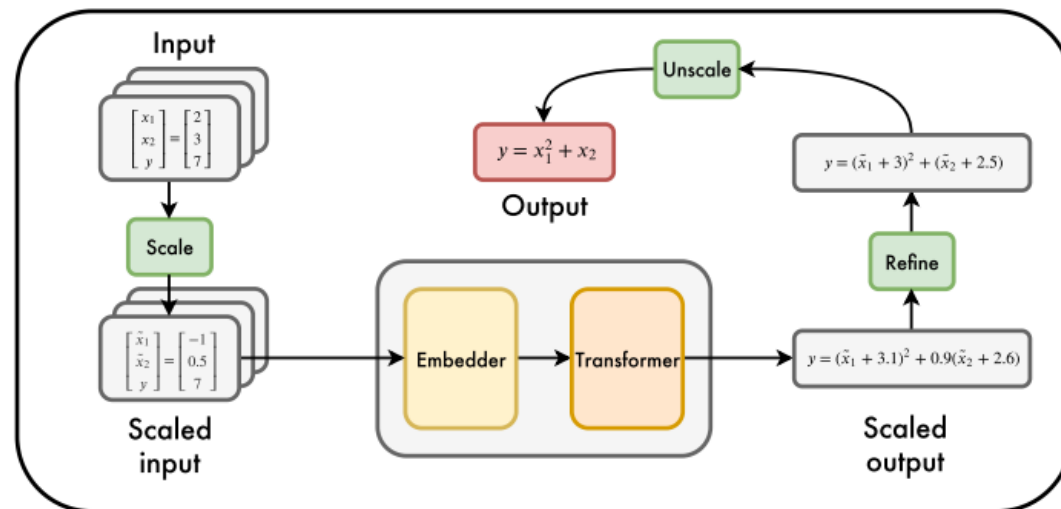


<https://doi.org/10.1557/mrc.2019.85>

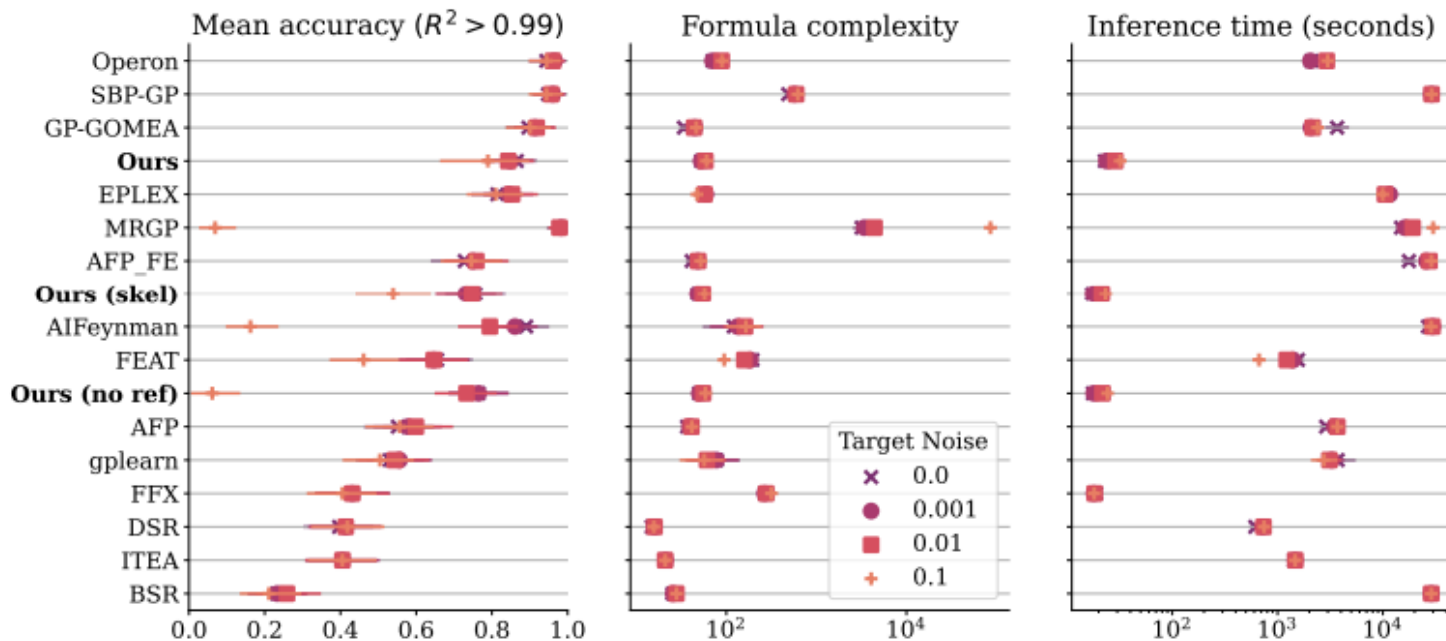
End-to-end symbolic regression with transformers



Training



Inference

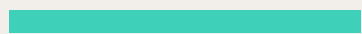


b_{\max}	Max binary ops	$5 + D$
O_b	Binary operators	add:1, sub:1, mul:1
u_{\max}	Max unary ops	5
O_u	Unary operators	inv:5, abs:1, sqr:3, sqrt:3, sin:1, cos:1, tan:0.2, atan:0.2, log:0.2, exp:1

[arXiv:2204.10532](https://arxiv.org/abs/2204.10532)

<https://github.com/facebookresearch/symbolicregression>

03



Экспериментальная часть

Описание эксперимента

Генерация модельных значений электронной плотности

- 1) от 0 до 1000 electrons per bohr³; 1500 и 1000 точек соответственно для обучения и тестирования модели
- 2) от 0 до 1e7 electrons per bohr³ (верхняя граница определялась расчетом электронной плотности атома Hg) – 2000 и 3000 точек соответственно для обучения и тестирования модели

Использование сгенерированных значений электронной плотности для расчета значений обменно-корреляционных потенциалов с известными аналитическими выражениями – *lda_x*, *lda_c_chachiyo*, *lda_c_pw_mod*

Расчет полной энергии молекулы воды в основном состоянии с использованием полученных аналитических формул и сравнение с референсным значением, полученным *ab initio*

Аппроксимация нейросетевого функционала неизвестного аналитического вида* моделями символьной регрессии и оценка полученной формулы

[*https://github.com/ml-electron-project/NNfunctional/tree/master](https://github.com/ml-electron-project/NNfunctional/tree/master)

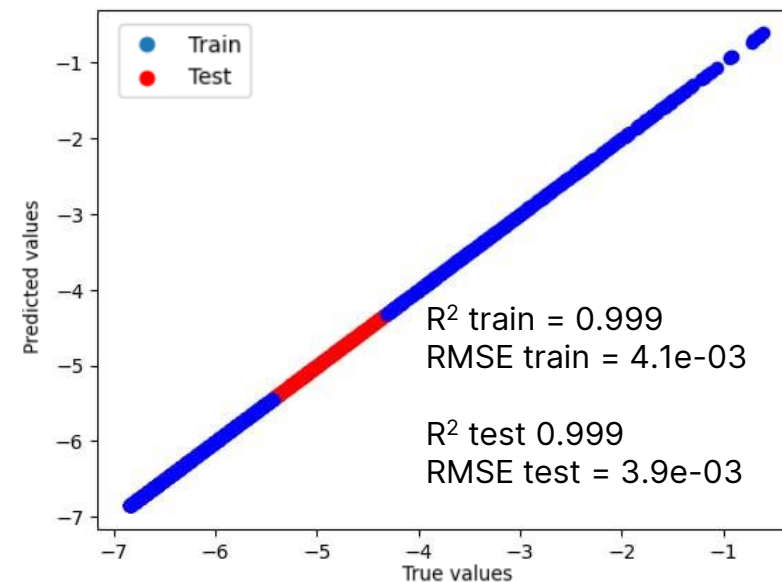
Символьная регрессия известных потенциалов генетическим алгоритмом

Интервал: от 0 до 1e7 electrons
per bohr³

Ida_x функционал

$$\left. \begin{aligned} & -\frac{3}{4} \left(\frac{3}{\pi} \right)^{\frac{1}{3}} n^{\frac{1}{3}} \\ & -0.7386 n^{\frac{1}{3}} \end{aligned} \right\} \text{референсная формула}$$

$$-0.738 n^{\frac{1}{3}} \quad \text{предсказанное выражение}$$

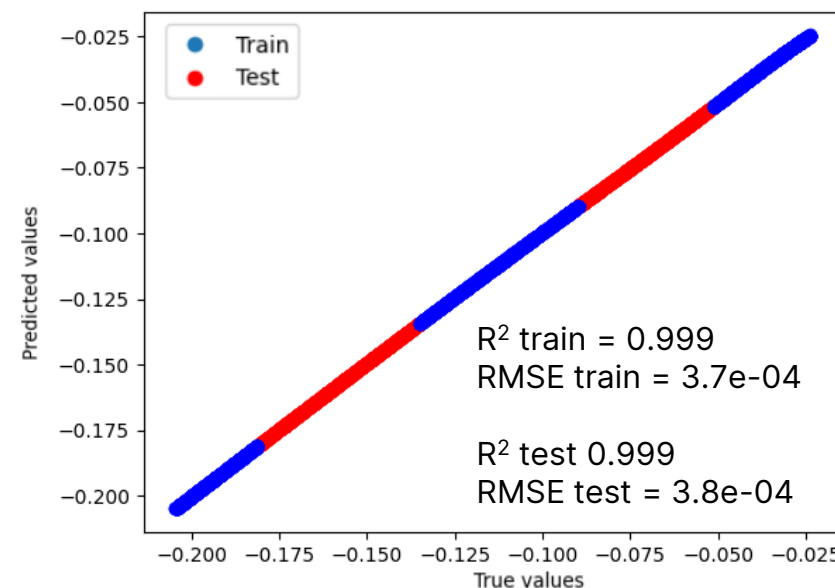


Интервал: от 0 до 1e7 electrons
per bohr³

Ida_c_chachiyo функционал

$$\left. \begin{aligned} & A * \ln \left(1 + b \left(\frac{4\pi}{3} \right)^{\frac{1}{3}} n^{\frac{1}{3}} + b \left(\frac{4\pi}{3} \right)^{\frac{2}{3}} n^{\frac{2}{3}} \right) \\ & -0.0156 * \ln(53.2 n^{\frac{2}{3}} + 33.0 n^{\frac{1}{3}} + 1) \end{aligned} \right\} \text{референсная формула}$$

$$-0.0219 * \ln \left(\left(n^{\frac{1}{3}} + \sqrt{(n + 0.167)} \right) - 0.0527 \right) \quad \text{предсказанное выражение}$$



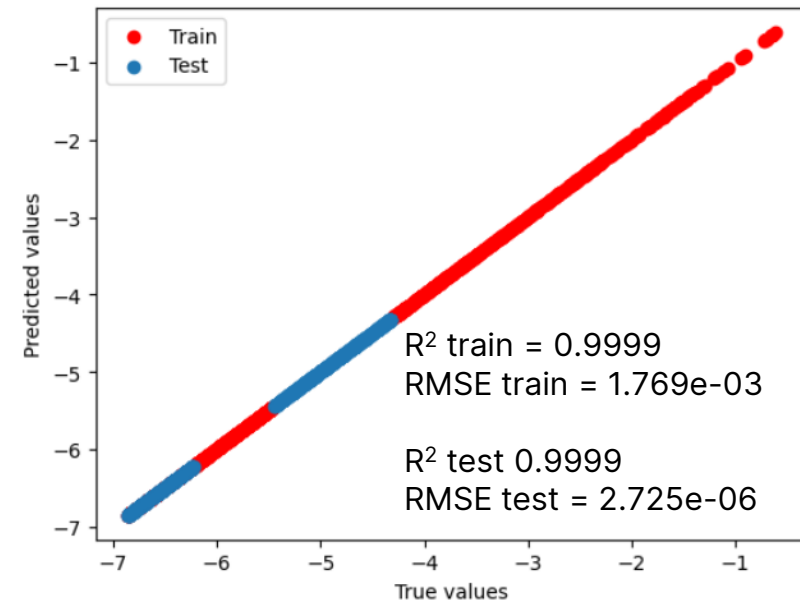
Символьная регрессия известных потенциалов нейросетью-трансформером

Интервал: от 0 до 1e4 electrons
per bohr³

Ida_x функционал

$$\left. \begin{aligned} &-\frac{3}{4} \left(\frac{3}{\pi} \right)^{\frac{1}{3}} n^{\frac{1}{3}} \\ &-0.7386 n^{\frac{1}{3}} \end{aligned} \right\} \text{референсная формула}$$

$$0.0027368415157811958 - 0.7386999562105924 n^{\frac{1}{3}} \quad \text{предсказанное выражение}$$



Интервал: от 0 до 1e4 electrons
per bohr³

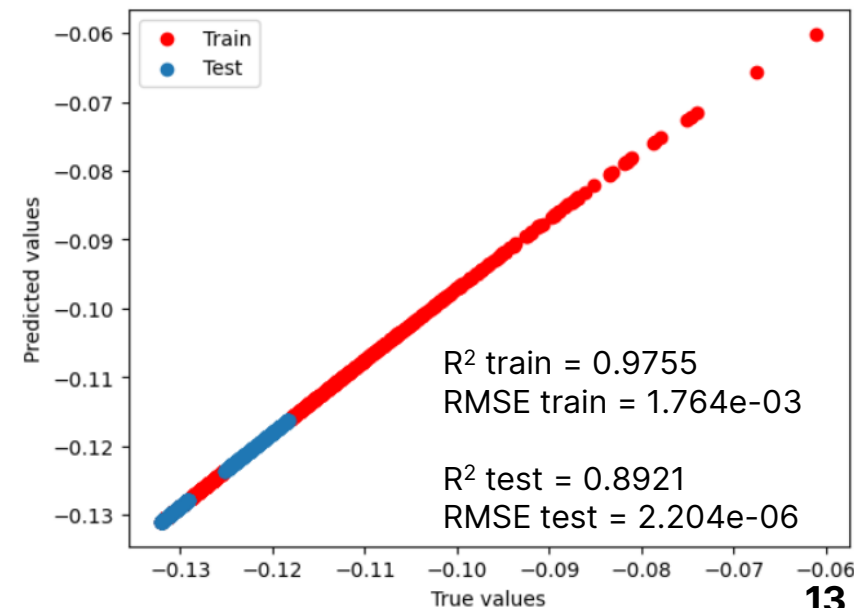
Ida_c_chachiyo функционал

$$\left. A * \ln \left(1 + b \left(\frac{4\pi}{3} \right)^{\frac{1}{3}} n^{\frac{1}{3}} + b \left(\frac{4\pi}{3} \right)^{\frac{2}{3}} n^{\frac{2}{3}} \right) \right\} \text{референсная формула}$$

$$-0.0156 * \ln(53.2 n^{\frac{2}{3}} + 33.0 n^{\frac{1}{3}} + 1)$$

$$-0.0172 * \ln(16.1 n^{\frac{2}{3}} + 17.7 n^{\frac{1}{3}} + 4.78) - 0.00463$$

предсказанное
выражение

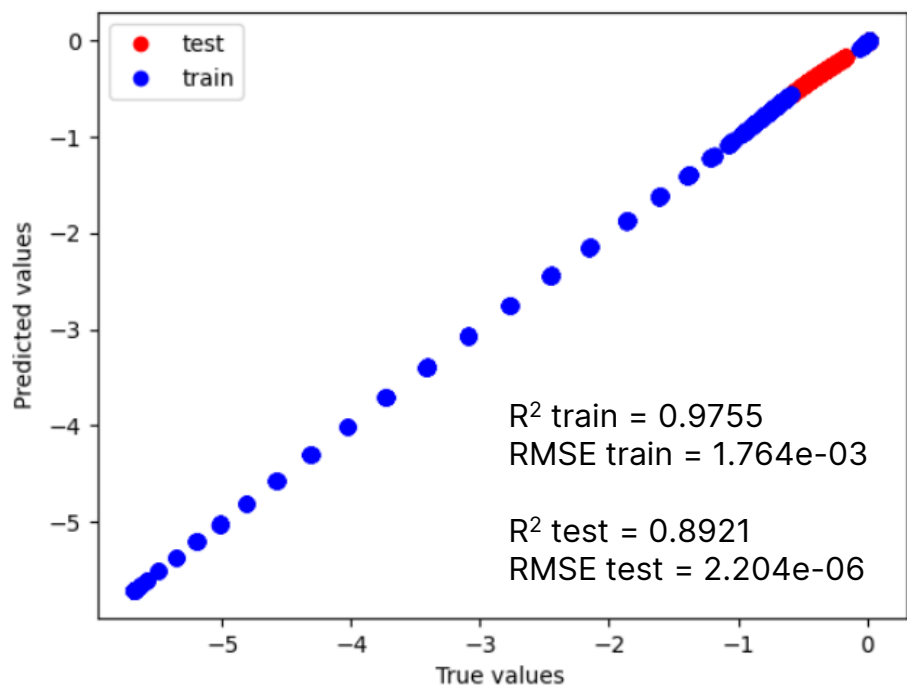


Символьная регрессия нейросетевого потенциала

Данные – рассчитанные ab initio значения электронной плотности системы и значения обменно-корреляционного потенциала в приближении LDA (lda_x + lda_c_chachiyo)

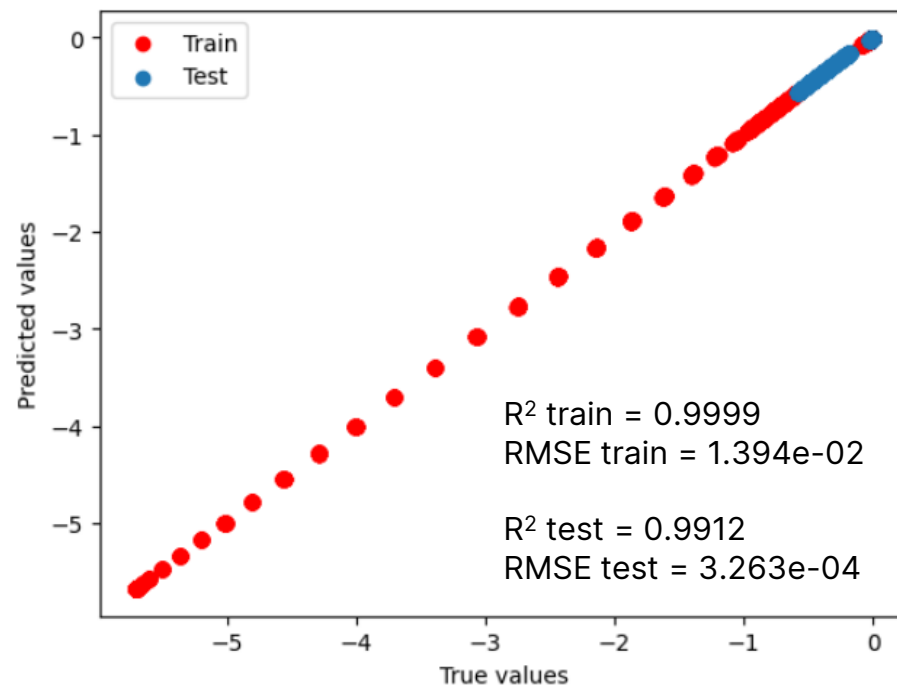
GA символьная регрессия

$$-0.9603n^{\frac{1}{3}} - 0.9603 * \left(\frac{1}{n + 0.6540} \right)^{1/6}$$



NN-трансформер символьная регрессия

$$(-18.95x - 0.2757)(0.0008950x + 0.03917)$$



Сравнение рассчитанных значений основного состояния молекулы воды

Используемый функционал	Значение энергии, в единицах Хартри	Абсолютная разность с референсным значением
Ab initio референсное значение*	-75.80	0
Нейросетевой функционал**	-75.83	0,03
GA символьная регрессия lda_x + lda_c_chachiyo	-61.87	13,93
GA символьная регрессия нейросетевого потенциала	-74.15	1,65
NN-трансформер символьная регрессия lda_x + lda_c_chachiyo	-74.07	1,73
NN-трансформер символьная регрессия нейросетевого потенциала	-73.90	1,9

*<https://pyscf.org/user/scf.html>

**<https://github.com/ml-electron-project/NNfunctional/tree/master>

04



Выводы



Выводы

- ✓ С помощью генетического алгоритма и нейросетевого трансформера были получены формульные выражения известных обменно-корреляционных функционалов в LDA приближении – Ida_x и $-da_c$ chachiyo. Полученные формулы показали высокое соответствие аналитическим зависимостям, а значения потенциала, посчитанные с помощью них, - соответствие потенциалам, посчитанным ab initio;
- ✓ Был аппроксимирован существующий нейросетевой обменно-корреляционный функционал с помощью моделей символьной регрессии. Посчитанные с помощью полученной формулы значения потенциалов показали соответствие значениям, полученным нейросетевым функционалом;
- ✓ Проведен расчет полной энергии основного состояния молекулы воды. Показано, что использование выражений, полученных с помощью метода символьной регрессии, позволяет рассчитывать энергию данной модельной системы с высокой точностью.

GitHub проекта

https://github.com/wwapper/AIRI_SR_exc_potentials.git



Наша команда



Ким
Павел,
ИТМО

Эксперименты с NN-трансформером для символьной регрессии



Кулаев
Кирилл,
ЮФУ

Эксперименты с генетическим алгоритмом для символьной регрессии



Рябов
Александр,
Сколтех

Формулирование гипотез, ведение проекта



Медведев
Михаил,
ИОХ РАН

Формулирование гипотез, ведение проекта



@AIRI_Research_Institute

Artificial Intelligence Research Institute

airi.net



[airi_research_institute](https://t.me/airi_research_institute)



[AIRI Institute](https://vk.com/AIRI_Institute)



[AIRI Institute](https://www.youtube.com/AIRI_Institute)



[AIRI_inst](https://twitter.com/AIRI_inst)



[artificial-intelligence-research-institute](https://www.linkedin.com/company/artificial-intelligence-research-institute)