

Глубокое обучение для обработки 3D структур белка

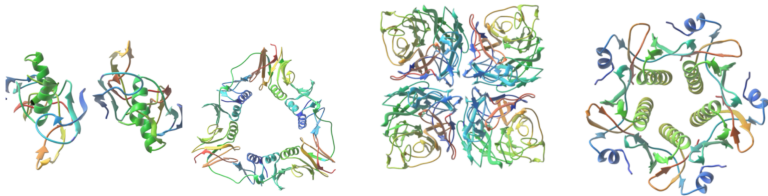
Веселова Е.Р.

Московский физико-технический институт

March 21, 2019

Проблематика

Обнаружение осей симметрии и повторяющихся элементов играет ключевую роль в исследовании белковых структур.



Примеры симметризованных белков

Основная литература

- Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data. CoRR, abs/1807.02547, 2018.
- Guillaume Pagès and Sergei Grudinin. DeepSymmetry: Using 3D convolutional networks for identification of tandem repeats and internal symmetries in protein structures. working paper or preprint, 2018.
- Georgy Derevyanko, Sergei Grudinin, Yoshua Bengio, and Guillaume Lamoureux. Deep convolutional networks for quality assessment of protein folds. ArXiv e-prints, 2018.

Постановка задачи

Данные

Даны $K = 11$ карт атомных плотностей элементов белка. Карты $x_i \in \mathbb{R}^{24 \times 24 \times 24}$ получены искусственной симметризацией датасета Top8000 с порядком циклической симметрии от 1 до $N_{\text{order}} \in [10, 20]$.

Модель

Исходная модель $f(X)$ — свёрточная нейросеть, $W \in \mathbb{R}^{n \times 4 \times 2}$ — матрица коэффициентов разложения каждого из 4 фильтров двух свёрточных слоёв нейросети на n базисных сферических гармоник.

Постановка задачи

Формулировка

Ответом на элементе выборки

$$f(x_i) = f_i \in \mathbb{R}^{N_{\text{order}}+6}$$

является композиция двух векторов

$$y_i \in \mathbb{R}^{N_{\text{order}}} \text{ и } z_i \in \mathbb{R}^6$$

где y_i определяет вероятность каждого порядка симметрии, z_i определяет положение оси симметрии.

Размерность 6 для задания оси симметрии выбрана неслучайно: модель переводит ось в трёхмерном пространстве в её представление в шестимерном пространстве посредством отображения Веронезе $V(x, y, z) = (x^2, y^2, z^2, \sqrt{2}yz, \sqrt{2}zx, \sqrt{2}xy)$.

Постановка задачи

Функции качества

$$L_c(W) = -\log(P(k_t)) = \log \left(\sum_{j=1}^{N_{\text{order}}} \exp(p_j) \right) - p_{k_t}.$$

$$L_a(W) = \|z_i - V(x_t, y_t, z_t)\|_2.$$

Функция потерь

$$W^* = \min_W (L_c(W) + L_a(W))$$

Постановка задачи

Особенность задачи

Требуется адаптировать модель $f(X, W)$ так, что при воздействии на исходные данные любым оператором $\pi(\text{tr})$ трансляции t и вращения r реализуется свойство эквивариантности, т.е.

$$f(\pi(\text{tr})X, W) = [\pi(\text{tr})f](X, W).$$

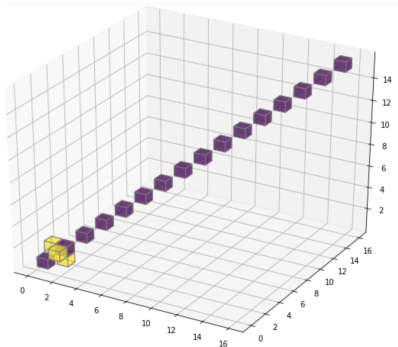
Эквивариантность реализуется при помощи разложения свёрточных фильтров в устойчивые к вращениям линейные комбинации сферических функций.

Базовый эксперимент

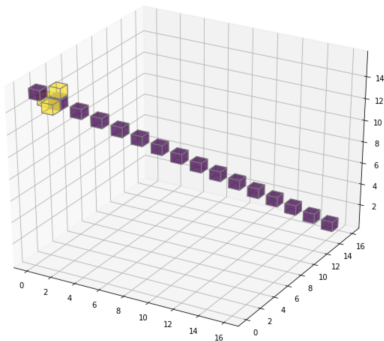
Воспроизведение структуры рабочей нейросети для выделения симметрий с помощью эквивариантных свёрточных слоёв.

Свёрточный слой взят из библиотеки se3cnn (Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data. CoRR, abs/1807.02547, 2018).

Результаты



[0.9348, 0.9518, 1.0000]



[-1.0940, 1.0106, 1.0426]

Дальнейшие исследования

Предлагается протестировать модель на реальных данных и найти оптимальную структуру.