

Предсказание свойств и типов атомов в молекулярных графах при помощи сверточных сетей

Манучарян Вардан

Кадукова Мария, Грудинин Сергей, Стрижов В.В.

Московский физико-технический институт

Курс: Численные методы обучения по прецедентам (практика, В.В.
Стрижов)/Группа 594, весна 2018

Цель исследования

- Цель работы

Предложить метод предсказания типов атомов по молекулярному графу

- Проблема

Существующие модели предсказывают тип на основе гибридизации с цепочкой условных операторов

- Метод решения

Автоматизация процесса с помощью свёрточных нейронных сетей, учитывающих окружение атома и 3d структуру молекулярных графов

Литература по предсказанию типов атомов

- Mathias Niepert. Learning Convolutional Neural Networks for Graphs, 2015 [использование cNN на графах]
- Steven Kearnes. Molecular Graph Convolutions: Moving Beyond Fingerprints, 2016 [самая близкая работа]
- Maria Kadukova, Sergei Grudinin. Knodle: A Support Vector Machines-Based Automatic Perception of Organic Molecules from 3D Coordinates. 2016 [алгоритм для сравнения]

Постановка задачи классификации атомов

Пусть $\mathfrak{S} = \{\mathfrak{s}_1, \dots, \mathfrak{s}_m\}$ - множество атомов в различных молекулах. Пусть $\mathbf{y} = \{y_1, \dots, y_m\}$ - типы атомов.

Пусть $G = \{g_1, \dots, g_n\}$ - набор функций, таких что $\forall i \forall j$ g_j отображает \mathfrak{s}_i в (i, j) элемент матрицы X :

$$g_j : (b_j, \mathfrak{s}_i) \rightarrow x_{ij} \in \mathbb{R}^1,$$

где b_j - набор параметров для g_j .

Определим модель f , сопоставляющую каждой строке X число из $[0, 1]$

$$f(w, X) = \frac{1}{1 + \exp(-Xw)},$$

где оптимальные параметры \hat{w} минимизируют функцию потерь

$$\hat{w} = \arg \min_w S(w|f, X, y),$$

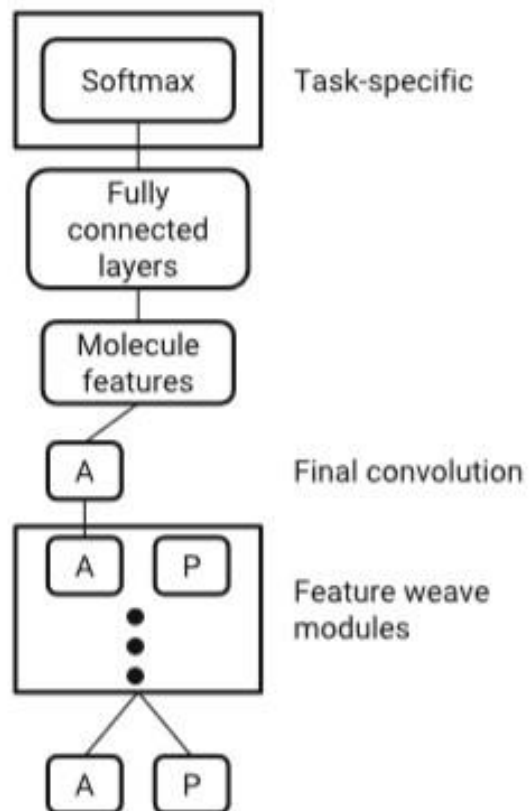
где

$$S(w|f, X, y) = -\ln\left(\sum_{i=1}^m y_i \log f(x_i, w) + (1 - y_i) \log(1 - f(x_i, w))\right)$$

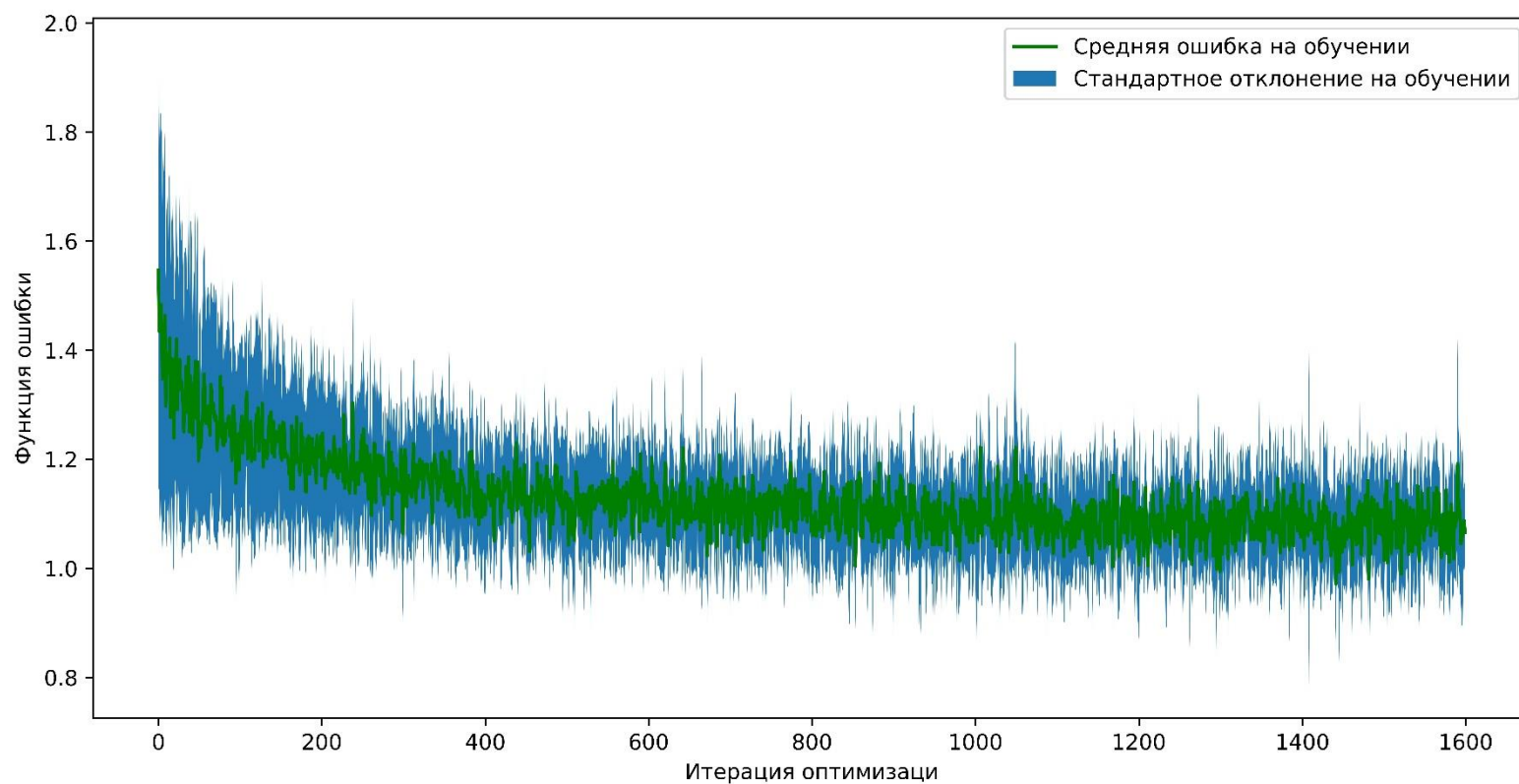
Цели эксперимента

- Получить результат, сопоставимый по качеству с Knodle
- Функционал качества – Accuracy

Эксперимент



Результаты эксперимента



Заключение

- Качество построенного алгоритма 45%
- Качество Knodle
- Необходимо улучшить модель: добавить признаки, изменить архитектуру