

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

*Манучарян Вардан*

manucharyan.va@phystec.edu

Московский физико-технический институт (государственный университет), Москва

Статья посвящена определению типов атомов и валентности в молекулярных графах при помощи методов машинного обучения. Предлагается использовать свёрточные нейронные сети (сNN), обученные на 3D структуре молекулярных графов. Для обучения используется  $\langle \dots \rangle$  датасет, в котором определены тип атома (около 150 классов), гибридизация атома (4 класса) и тип связи (5 классов). Результат работы нового алгоритма будет сравниваться с Knodle [1].

**Ключевые слова:** машинное обучение, классификация, свёрточные нейронные сети, молекулярные графы.

## 1 Введение

### 1.1

В этой работе мы хотим научиться эффективно классифицировать атомы, для того чтобы предсказывать взаимодействие молекул. Это важно во многих вычислительных методах в медицине и биологии, например, виртуальный скрининг при разработке новых лекарств [2]. При этом молекулы представимы в виде трёхмерных молекулярных графов, что позволяет использовать методы машинного обучения на графах.

“В последние годы были предложены несколько алгоритмов. Самые ранние работы основывались на простых геометрических соображениях, использующих длины связей и валентные углы [3]. Позже стали учитывать функциональные группы [4], гибридизацию и заряд атома [5–8]. Чтобы уменьшить влияние ошибок в экспериментальном определении структуры были использованы некоторые подходы: максимальное взвешенное паросочетание [9, 10], поиск структуры Льюиса [11], максимизация [12] или минимизация [13, 14] некой метрики.”

Также используют модели, основанные на свёрточных нейронных сетях (сNN) на графах [15, 16].

В данной работе предлагается  $\langle \dots \rangle$ .

Сравниваться данный алгоритм будет с алгоритмом, реализованным в библиотеке по распознаванию типов атомов Knodle [1], основанным на мультиклассовой классификации при помощи метода опорных векторов.

## 2 Постановка задачи

Будем использовать файлы формата mol2 из базы данных PDBBindCN.

## Литература

- [1] *Maria Kadukova, Sergei Grudinin* Knodle: A Support Vector Machines-Based Automatic Perception of Organic Molecules from 3D Coordinates // Journal of Chemical Information and Modeling, American Chemical Society, 2016, 56 (8), pp.1410-1419.

\*Научный руководитель: Стрижов В.В. Задачу поставил: Стрижов В.В. Консультанты: Грудинин С., Кадукова М.

- [2] *Bohdan Waszkowycz, David E Clark, and Emanuela Gancia* Outstanding challenges in protein–ligand docking and structure-based virtual screening // *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 1(2):229–259, 2011.
- [3] *Jon C Baber and Edward E Hodgkin* Automatic assignment of chemical connectivity to organic in the Cambridge structural database // *J. Chem. Inf. Comput. Sci.*, 32(5):401–406, 1992.
- [4] *Manfred Hendlich, Friedrich Rippmann, and Gerhard Barnickel* Bali: Automatic assignment of bond and atom types for protein ligands in the brookhaven protein databank // *J. Chem. Inf. Comput. Sci.*, 37(4):774–778, 1997.
- [5] *Elke Lang, Claus-Wilhelm von der Lieth, and Thomas Forster* Automatic assignment of bond orders based on the analysis of the internal coordinates of molecular structures // *Anal. Chim. Acta*, 265(2):283–289, 1992.
- [6] *Yuan Zhao, Tiejun Cheng, and Renxiao Wang* Automatic perception of organic molecules based on essential structural information // *J. Chem. Inf. Model.*, 47(4):1379–1385, 2007.
- [7] *Daan MF van Aalten, R Bywater, John BC Findlay, Manfred Hendlich, Rob WW Hooft, and Gert Vriend* Prodrgr, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules // *J. Comput.-Aided Mol. Des.*, 10(3):255–262, 1996.
- [8] *Qian Zhang, Wei Zhang, Youyong Li, Junmei Wang, Liling Zhang, and Tingjun Hou* A rule based algorithm for automatic bond type perception // *J. Cheminf.*, 4(1):1–10, 2012.
- [9] *Paul Labute* On the perception of molecules from 3d atomic coordinates // *J. Chem. Inf. Model.*, 45(2):215–221, 2005.
- [10] *Gerd Neudert and Gerhard Klebe* Fconv: Format conversion, manipulation and feature computation of molecular data // *Bioinformatics*, 27(7):1021–1022, 2011.
- [11] *Matheus Froeyen and Piet Herdewijn* Correct bond order assignment in a molecular framework using integer linear programming with application to molecules where only non-hydrogen atom coordinates are available // *J. Chem. Inf. Model.*, 45(5):1267–1274, 2005.
- [12] *Sascha Urbaczek, Adrian Kolodzik, Inken Groth, Stefan Heuser, and Matthias Rarey* Reading pdb: Perception of molecules from 3d atomic coordinates // *J. Chem. Inf. Model.*, 53(1):76–87, 2012.
- [13] *Junmei Wang, Wei Wang, Peter A Kollman, and David A Case* Automatic atom type and bond type perception in molecular mechanical calculations // *J. Mol. Graphics Modell.*, 25(2):247–260, 2006.
- [14] *Anna Katharina Dehof, Alexander Rurainski, Quang Bao Anh Bui, Sebastian Bocker, Hans-Peter Lenhof, and Andreas Hildebrandt* Automated bond order assignment as an optimization problem // *Bioinformatics*, 27(5):619–625, 2011.
- [15] *Mathias Niepert, Mohamed Ahmed, Konstantin Kutzkov* Learning Convolutional Neural Networks for Graphs // , 2016
- [16] *Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, Patrick Riley* Molecular Graph Convolutions: Moving Beyond Fingerprints // , 2016

*Поступила в редакцию*