

Rozšírené zadanie diplomovej práce

Samuel Baran

Názov práce: Nové techniky učenia sa bez učiteľa pre klasifikáciu a predikciu molekulárnych vlastností
Zadávatel': Tachyum s.r.o.
Vedúci práce (Tachyum): Ing. Slavomír Maťašovský
Vedúci práce (UPJŠ): RNDr. Ľubomír Antoni, PhD.

Úvod a motivácia

Strojové učenie je podmnožinou umelej inteligencie, pričom sa zaoberá metódami a algoritmi učenia sa stroja z údajov. Cieľom strojového učenia je modelovanie algoritmov učenia sa pomocou stroja na základe vstupných dát v definovanom priestore riešení. Strojové učenie bolo v poslednej dobe úspešne aplikované na riešenie úloh v mnohých oblastiach vedy. V oblasti predikcie a klasifikácie molekulárnych vlastností sa stretávame s nižšou dostupnosťou údajov a vyššou heterogenosťou vstupných údajov. Tieto skutočnosti vplyvajú na nevhodnosť používania metód kontrolovaného učenia v tejto oblasti. Z týchto dôvodov vzniká priestor pre využitie metód učenia sa bez učiteľa, ktorého prednosťou je primárne využitie neoznačených dát na vytvorenie skrytých reprezentácií vstupov, ktoré poskytujú lepšiu východiskovú pozíciu pre modely učenia sa s učiteľom.

Techniky učenia sa bez učiteľa

Nové techniky učenia sa bez učiteľa alebo čiastočného učenia bez učiteľa (t. j. *semi-supervised learning*) s využitím rôznych typov neurónových sietí boli úspešne použité najmä v oblasti počítačového videnia. K novým metódam v tejto oblasti patrí napríklad tzv. samo-destilácia bez použitia značiek (*self-distillation with no labels*), ktorá bola vhodne použitá pri problémoch počítačového videnia. Možným smerom v oblasti molekulárnych vlastností je adaptácia a testovanie s využitím modelov pre klasifikáciu a predikciu vlastností molekúl a porovnanie ich efektivity s doposiaľ najúspešnejšími technikami (napríklad s konfrontačným učením), ktoré boli vyvinuté priamo pre túto oblasť.

Ciele práce

Pre túto bakalársku prácu boli sformulované nasledujúce ciele:

- Spracovať prehľad metód klasifikácie a predikcie vlastností molekúl a najnovších techník učenia sa bez učiteľa alebo čiastočného učenia bez učiteľa, ktoré sú využívané v oblasti počítačového videnia.
- Navrhnuť novú metódu klasifikácie a predikcie vlastností molekúl, ktorá bude inšpirovaná existujúcimi technikami učenia sa bez učiteľa alebo čiastočného učenia bez učiteľa z oblasti počítačového videnia.
- Implementovať a experimentálne overiť navrhnuté riešenie testovaním na dátach pre klasifikáciu a predikciu molekulárnych vlastností.
- Porovnať dosiahnuté výsledky navrhnutého riešenia s doteraz najúspešnejšími technikami vyvinutými v oblasti molekulárnych vlastností.

Odporúčaná literatúra:

1. Zhenqin Wu and Bharath Ramsundar and Evan N. Feinberg and Joseph Gomes and Caleb Geniesse and Aneesh S. Pappu and Karl Leswing and Vijay S. Pande, . "MoleculeNet: A Benchmark for Molecular Machine Learning". CoRR abs/1703.00564. (2017).
2. DeepChem - <https://deepchem.readthedocs.io/en/latest/index.html>
3. Mathilde Caron and Hugo Touvron and Ishan Misra and Hervé Jégou and Julien Mairal and Piotr Bojanowski and Armand Joulin, . "Emerging Properties in Self-Supervised Vision Transformers". CoRR abs/2104.14294. (2021).
4. Mahmoud Assran and Mathilde Caron and Ishan Misra and Piotr Bojanowski and Armand Joulin and Nicolas Ballas and Michael G. Rabbat, . "Semi-Supervised Learning of Visual Features by Non-Parametrically Predicting View Assignments with Support Samples". CoRR abs/2104.13963. (2021).
5. Yuyang Wang and Jianren Wang and Zhonglin Cao and Amir Barati Farimani, . "MolCLR: Molecular Contrastive Learning of Representations via Graph Neural Networks". CoRR abs/2102.10056. (2021).
6. Seyone Chithrananda and Gabriel Grand and Bharath Ramsundar, . "ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction". CoRR abs/2010.09885. (2020).