

# Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

**Autor práce** Adam Frey  
**Název práce** Predicting novel drug-target interactions via deep learning techniques  
**Rok odevzdání** 2020  
**Studijní program** Informatika    **Studijní obor** Umělá inteligence

**Autor posudku** Miroslav Kratochvíl    **Role** oponent  
**Pracoviště** Katedra softwarového inženýrství

## Text posudku:

Cílem práce je prozkoumat aplikaci populárních metod strojového a hlubokého učení na farmakologický problém rozšiřování aplikovatelnosti léčiv (“drug repurposing”) a výsledky potvrdit experimentem; práce tento cíl splňuje.

Hlavním autorovým přínosem a výsledkem je aplikace různých zlepšení na stávající modely učení. V porovnání se publikovaným výzkumem (2018) se autorovi podařilo zvýšit kvalitu predikce o několik (cenných) procent v metrikách AuC/AuPR, a navíc poukázat na prozatím neidentifikované, nadějně vypadající kandidáty pro drug repurposing. Výsledek experimentu (tabulky 4.1 a 4.2) se celkově dá považovat za úspěch.

Implementace experimentu je provedená v Pythonu způsobem obvyklým pro framework TensorFlow, kód má celkem asi 500 řádek, drobná programátorská dokumentace je k dispozici. Použitý dataset je z velké části převzatý z publikace. U autorova příspěvku ke vstupním datům (deskriptorům/fingerprintům sloučenin z databáze DrugBank) bych očekával nějakou reprodukovatelnou metodu (např. skript) sloužící k vygenerování z jasně identifikovaných základních dat, v tomto případě nejspíš z DrugBankového SDF. Metodologicky by se dalo očekávat porovnávání více kombinací parametrů modelu; například by bylo vhodné ověřit, jestli model využívající chemické deskriptory (popsaný v sekci 4.4) nepůjde ‘opravit’ většími latentními vektory nebo lepšími deskriptory (PubChem fingerprint byl navržený pro akceleraci vyhledávání v databázích, pro drug repurposing by možná byly vhodnější deskriptory používané v QSAR modelech, typicky ECFP).

Většinu textové části práce tvoří poměrně rozsáhlý popis základních technik strojového učení a hrubý přehled terminologie a technik ve výzkumu interakcí proteinů a malých molekul. Ten se dá považovat za poměrně dobrý úvod do problematiky, samotným výsledkům a autorově přínosu je pak v textu bohužel věnováno poměrně málo místa.

Text práce trpí následujícími nedostatky:

- Autorovy výsledky a návrh experimentu by mohly být popsány přehledněji, ideálně s předstihem před technickými detaily, případně naznačeny už v úvodu práce, který se teď konkrétních cílů dotýká jen velmi okrajově. Motivace pro většinu textu práce by pak byla jasnější.

- Popis výsledků je sice dostatečný, je ale dobrým zvykem podobné experimenty prezentovat systematictěji jako hypotézy, a tyto následně potvrzovat nebo vyvracet empiricky testováním. Přehledný seznam dokázaných nebo vyvrácených hypotéz chybí.
- Text obsahuje několik špatně pochopitelných souvětí (mimo jiné např. popis grafového vstupu neuronových sítí v sekci 3.4.3 strana 50, nebo poněkud matoucí motivaci pro aproximaci výpočtu chyb hypotézy v sekci 3.1.1, strana 19, třetí odstavec). Vyskytují se občasné problémy se pořádkem slov a členy (např. ‘an another’ str. 31 odst. 4, nebo několikrát se vyskytující ‘An example is. . .’), mírné nekonzistence (např. nenápadné ‘grows smaller’, str. 43 dole), a různé další drobnosti (např. apostrofové zkratky sloves).
- Úvod do problematiky je poměrně obsáhlý až zdouhavý, naproti tomu popisu hlavních algoritmů použitých přímo pro experiment (např. low-rank matrix factorization) a jejich napojení na zbytek hlubokého učení je věnováno poměrně málo místa. Některé dlouhé bloky nestrukturovaného textu (sekce 2.2 a 3.2.1–3.2.3) by mohly být rozděleny nebo systematizovány, např. pomocí různých podsekcí nebo seznamů.
- Občas se vyskytují nepříjemné formátovací potíže, např. nadpis sekce 3.4.2, rozlišení některých grafů (obrázky 2.4–2.7), poněkud nestandardní formát citačních referencí s dvojitými hranatými závorkami, chybějící odkaz na (trochu nepřátelsky barevné) obrázky 2.1 a 2.2 v textu, a nevysvětlené číslování v obr. 2.1.

Předpokládám, že hlavní námitky k práci by bylo možné odstranit dostatečným jazykovým a strukturálním editováním textové části. Celkově by proto neměly tvořit překážku obhajitelnosti jinak kvalitních výsledků práce.

**Práci doporučuji k obhajobě.**

**Práci nenavrhuji na zvláštní ocenění.**

V Praze dne 26. 1. 2020

Podpis: