



Posudek oponenta závěrečné práce

Oponent práce: Ing. Jitka Hrabáková
Student: Roman Bushuiev
Název práce: Predikce biosyntézy terpenů pomocí strojového učení
Obor / specializace: Znalostní inženýrství
Vytvořeno dne: 4. června 2021

Hodnotící kritéria

1. Splnění zadání

- ▶ [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Zadání předložené bakalářské práce je zcela splněno. Samotné téma považuji za obtížné, neboť nad rámec problematiky strojového učení vyžaduje porozumění chemickým zápisům molekulární struktury a její reprezentaci vhodné pro strojové učení.

2. Písemná část práce 100/100 (A)

Rozsah písemné práce je v pořádku, neobsahuje zbytečné části a přehledně popisuje zkoumanou problematiku. Text je psán v angličtině téměř bez chyb, je dobře strukturovaný a přehledný. Student cituje relevantní zdroje. Student jasně odděluje vlastní práci od převzatých výsledků. Práce je po formální stránce zcela v pořádku.

3. Nepísemná část, přílohy 90/100 (A)

Student vytvořil implementaci pipeline tří modelů strojového učení kód je přehledně organizován do poměrně velkého množství Jupyter Notebooku.

4. Hodnocení výsledků, jejich využitelnost 95/100 (A)

Práce vznikala jako součást projektu na Ústavu Organické Chemie a Biochemie, kde budou její výsledky pravděpodobně dále rozvíjeny a používány.

Celkové hodnocení

95 /100 (A)

Práci považuji za zdařilý první krok k aplikaci metod strojového učení v predikci biosyntézy terpenů. Student prozkoumal užití neuronových sítí v jiných odvětvích biochemie a aplikoval je na zkoumaný problém. Navrhl pipeline tří modelů strojového učení, nastavil hyperparametry, provedl fine-tuning přetrénovaných modulů a natrénoval MPL modul. Výsledný model otestoval na testovací množině a výsledky přehledně prezentuje a komentuje. Oceňuji tvůrčí schopnosti studenta, např. vlastní volba metrik pro určení vhodného fingerprint molekuly, vypořádání se s malou množinou trénovacích dat, možnost vylepšení výsledků pomocí nevyužité informace o chemickém vzorci a další.

Otázky k obhajobě

1. V tabulce 5.1 validační skóre pro testované modely pozorujeme následující. Pro MLP 1L se s narůstajícím počtem trénovacích epoch značně zvyšuje počet nedekódovaných latentních vektorů (#0), ačkoli ostatní skóre zůstávají téměř stejná, podobné chování ovšem nepozorujeme u MLP 3L. Je pro tento jev vysvětlení?
2. Navrhujete mnoho zajímavých možností pro další práci. Máte v úmyslu v započaté práci pokračovat?

Instrukce

Splnění zadání

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

Nepísemná část, přílohy

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

Hodnocení výsledků, jejich využitelnost

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Celkové hodnocení

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.