

Posudek oponenta diplomové práce

Student: Kohout Pavel, Bc.
Téma: Využití variačních autoenkodérů pro ancestrální rekonstrukci sekvencí (id 24721)
Oponent: Martínek Tomáš, Ing., Ph.D., UPSY FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Cílem práce bylo aplikovat techniku strojového učení na bázi variačních autoenkodérů na problém ancestrální rekonstrukce proteinových sekvencí. Následně bylo potřeba vyhodnotit dosažené výsledky a porovnat je s již publikovanými výstupy. **Zadání považuji za obtížnější.**
- 2. Splnění požadavků zadání** **zadání splněno s podstatným rozšířením**
Nad rámec původního zadání student prozkoumal řadu metod pro prohledávání latentního prostoru za účelem nalezení věrohodných ancestrálních sekvencí. Celkově svou práci student doplnit řadou promyšlených experimentů a povýšil ji na úroveň vědecké práce.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
- 4. Prezentací úroveň předložené práce** **80 b. (B)**
Diplomová práce je napsána přehledně a kapitoly jsou uspořádány v logickém sledu. **Rozsah teoretické a praktické části je vyvážený.** Samotný text je pro čtenáře čitelný a pochopitelný. **Výhrady** bych měl pouze k některým **teoretickým částem obsahujícím řadu matematických vzorců** bez potřebného vysvětlení.
- 5. Formální úprava technické zprávy** **95 b. (A)**
Předložená **práce je na velmi dobré jazykové úrovni.** Rovněž typografická stránka textu je na vysoké úrovni. Text je doplněn názornými obrázky a grafy.
- 6. Práce s literaturou** **100 b. (A)**
Práce s literaturou je na velmi dobré úrovni. **Čerpáno bylo z kvalitních časopiseckých publikací** z oblasti molekulární biologie, bioinformatiky a ancestrální rekonstrukce. Převzaté části textu a obrázky jsou řádně označeny a odděleny od vlastního přínosu.
- 7. Realizační výstup** **100 b. (A)**
Hlavní realizační **výstupy práce tvoří jádro aplikace pro inferenci ancestrálních sekvencí** vytvořené v **jazyce python** a doprovodné skripty s experimenty a jejich spuštění v prostředí Metacentra. Uvedené **zdrojové kódy jsou plně funkční** a v souladu s licenčními podmínkami.
- 8. Využitelnost výsledků**
Předložená práce navrhuje **nový přístup** pro inferenci ancestrálních sekvencí **s využitím variačních autoenkodérů.** Přestože se může zdát, že tento nový přístup prozatím nedosahuje kvalit konvenčních metod, mají výstupy této práce nepochybně **publikační potenciál** v kvalitním vědeckém časopise.
- 9. Otázky k obhajobě**
 - Z textu mi nebylo zřejmé, zda mapování ASR sekvencí do latentního prostoru dopadlo dobře či nikoliv. Z obrázku 8.4 spíše plyne, že evoluční vztahy se v latentním prostoru nedaří vždy zachytit v dostatečné míře. Jak to tedy je?
 - Jak úspěšná byla rekonstrukce trénovacích sekvencí z latentního prostoru zpět do sekvenčního? Do jaké míry může tato rekonstrukce ovlivnit kvalitu generovaných ASR sekvencí?
 - Kolik trénovacích vzorků a jaké parametry enkodéru jsou obvyklé v úspěšných aplikacích této metody (např. v oblasti zpracování videa, zvuku)? Odpovídá to parametrům úloh v oblasti proteinového inženýrství?
- 10. Souhrnné hodnocení** **98 b. výborně (A)**
Předložená diplomová práce je velmi kvalitní a student prokázal, že je schopen samostatně pochopit poměrně náročnou problematiku a vytvořit aplikaci pro inferenci ancestrálních sekvencí s využitím variačních autoenkodérů. Rád bych ocenil velmi precizní přístup k analýze daného problému a metodickém zpracování, včetně řady vědeckých experimentů, které jsou na velmi vysoké úrovni. Vytvořený výstup má navíc potenciál se stát součástí kvalitní vědecké publikace. S ohledem na tyto skutečnosti hodnotím stupněm **výborně (A)** a doporučuji komisi zvážit případné ocenění této práce.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 1. června 2022

Martínek Tomáš, Ing., Ph.D.
oponent