

Posudek oponenta diplomové práce

Student: Grešová Katarína, Bc.

Téma: Klasifikace bakterií do taxonomických kategorií na základě vlastností 16s rRNA (id 22671)

Oponent: Hon Jiří, Ing., UIFS FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Cílem práce bylo predikovat poměry známých bakterií ve vstupním metagenomickém vzorku podle počtu jednotlivých variant genu 16S rRNA. Zadání považuji za obtížnější kvůli nutnosti nastudovat množství biologických souvislostí.
- 2. Splnění požadavků zadání** **student se odůvodněně odchýlil od zadání**
Téma se mírně odchýlilo od původního zadání směrem ke složitější úloze. Místo klasifikace jednotlivých genů 16S rRNA do taxonomických kategorií řešila studentka odhad množství známých bakterií v metagenomickém vzorku. Charakter zadání se tedy změnil z klasické klasifikační úlohy, kde na vstupu je jeden gen a na výstupu taxonomické zařazení, na optimalizační problém, jak přiřadit množinu genů ke známým bakteriím, aby dané přiřazení co nejlépe odpovídalo reálnému složení.
- 3. Rozsah technické zprávy** **je v obvyklém rozmezí**
- 4. Prezentací úroveň předložené práce** **75 b. (C)**
Předložená práce je celkově napsána přehledně a kapitoly jsou uspořádány v logickém sledu. Rozsah teoretické a praktické části je vyvážený. Samotný text do značné míry předpokládá znalost dané oblasti. K prezentační úrovni mám několik připomínek. Z textu není jasné, jak se přesně počítá kombinace různých evaluátorů, to lze dohledat až ve zdrojovém kódu. V textu nejsou zmíněny žádné existující přístupy. Pokud je použitý přístup zcela nový, mělo by to explicitně zaznít. Na mnoha místech v textu práce mi chybělo podrobnější vysvětlení motivace pro použité kroky a metody.
- 5. Formální úprava technické zprávy** **90 b. (A)**
Občas se objevují překlepy. Stylistickou úroveň nemohu kvalitně posoudit, jelikož je práce psána slovensky. Typografická stránka je kvalitní s výjimkou způsobu zápisu a formátování některých vzorců. Vzorce by si zasloužily doplnit číslováním. Text je doplněn názornými obrázky, schémata a grafy.
- 6. Práce s literaturou** **95 b. (A)**
Práce s literaturou je na dobré úrovni. Čerpáno bylo z kvalitních časopiseckých publikací z oblasti analýzy metagenomických dat. Převzaté části textu a obrázky jsou řádně označeny a odděleny od vlastního přínosu.
- 7. Realizační výstup** **75 b. (C)**
Hlavním výstupem práce jsou zdrojové kódy v jazyce Python, které umožňují zpracování datové sady, nastavení různých parametrů a následně prohledávání prostoru možných řešení. Dalším výstupem jsou experimenty s různými variantami nastavení a jejich zhodnocení. Rozsah zdrojových kódů není velký, což je pochopitelné vzhledem k experimentálnímu zaměření práce. Schází však více podrobnějších experimentů na větších datových sadách pro kvalitnější zhodnocení navrhované metody.
- 8. Využitelnost výsledků**
Práce řeší velice aktuální téma analýzy metagenomických vzorků. Prezentovaný přístup se novým způsobem snaží určovat druhy a poměry jednotlivých bakterií v metagenomickém vzorku. Po dalším dopracování a pečlivějším vyhodnocení by výsledky mohly být publikovány v odborném časopise.
- 9. Otázky k obhajobě**
 1. V práci nejsou uvedeny žádné existující nástroje pro danou úlohu. Existují nějaké?
 2. Jak bude ovlivněn výstup vašeho algoritmu, pokud bude vstup obsahovat geny neznámých bakterií?
 3. Mohla byste použít i jinou metodu pro prohledávání stavového prostoru, než je prezentovaný algoritmus Metropolis Hastings? Jaké by to případně mělo výhody a nevýhody?
- 10. Souhrnné hodnocení** **85 b. velmi dobře (B)**
Studentka prokázala, že je schopna se orientovat v obtížnější problematice analýzy metagenomických dat. Navrhla a implementovala novou metodu pro predikci bakteriálního složení metagenomického vzorku. Oceňuji, že studentka implementovala a ověřila mnoho způsobů ohodnocení kvality predikovaného přiřazení a jejich kombinace. Celkové zhodnocení navrhované metody by však mělo být podrobnější a především by mělo být provedeno na rozsáhlejších reálných datech. Vzhledem k chybějícímu podrobnému zhodnocení a spíše průměrné prezentační úrovni práce navrhuji hodnocení stupněm **velmi dobře (B)**.

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

V Brně dne: 25. června 2020

Hon Jiří, Ing.
oponent