

**I. IDENTIFIKAČNÍ ÚDAJE**

|                                   |   |
|-----------------------------------|---|
| <b>Název práce:</b>               | Haplotype Estimation Using Sequencing Reads |
| <b>Jméno autora:</b>              | bc. Anastasia Lebedeva                      |
| <b>Typ práce:</b>                 | díplomová                                   |
| <b>Fakulta/ústav:</b>             | Fakulta elektrotechnická (FEL)              |
| <b>Katedra/ústav:</b>             | Katedra počítačů                            |
| <b>Oponent práce:</b>             | Doc. Ing. Jiří Kléma, PhD.                  |
| <b>Pracoviště oponenta práce:</b> | Katedra počítačů FEL ČVUT                   |

**II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ**

|   |                   |
|---|-------------------|
| <b>Zadání</b>   | <b>náročnější</b> |
| <i>Hodnocení náročnosti zadání závěrečné práce.</i>   |                   |
| Zadání se zaměřuje na rozšíření současného populárního algoritmu populačního genotypového fázování o fázování fragmentové. Úkolem diplomantky bylo při odhadu haplotypu využít jako doplňkovou informaci ze sekvenování ve formě fragmentů dlouhých úseků DNA odpovídajících fragmentům haplotypu. Zadání diplomové práce předpokládá dobrou znalost genetiky i informatiky. Diplomantka musela nastudovat základní genetické minimum, seznámit se s existujícími netriviálními fázovacími algoritmy, formáty vstupních dat, navrhnout algoritmus řešení a následně jej implementovat a otestovat. Z všech těchto důvodů zadání pokládám za spíše náročnější. |                   |

|   |                                    |
|---|------------------------------------|
| <b>Splnění zadání</b>   | <b>splněno s menšími výhradami</b> |
| <i>Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>  |                                    |
| Diplomantka hlavní cíle práce splnila. Porovnávám-li zadání s řešením nacházím menší rozdíl v zamýšlené možné aplikaci pro genotypové refinování a zadání také předpokládalo srovnání algoritmu s jinými populačními a fragmentovými algoritmy. V práci vidím porovnání různých parametrizací současného řešení, tj. de facto srovnání výchozího algoritmu s modifikovaným. |                                    |

|   |                |
|---|----------------|
| <b>Zvolený postup řešení</b>  | <b>správný</b> |
| <i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>   |                |
| Zvolený postup řešení pokládám za správný. Experimenty v závěru práce naznačují, že se podařilo zvýšit přesnost fázování za cenu rozumného navýšení doby výpočtu. Přesnost fázování podle očekávání roste s délkou fragmentů a klesá s jejich chybovostí. Očekávaný je také vliv parametru pokrytí při sekvenování. |                |

|  |                    |
|--|--------------------|
| <b>Odborná úroveň</b>  | <b>A - výborně</b> |
| <i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>   |                    |
| Nenarazil jsem na žádné odborné nedostatky. Diplomantka volila vhodný výchozí algoritmus, pracuje se správnými vstupními daty (panely čtení apod.) a datovými formáty. I když jsem v omezeném čase nebyl schopen ověřit korektnost pravděpodobnostní inferencí haplotypů (viz otázka na konci posudku) odbornou úroveň odvozuji i od dosažených experimentálních výsledků. |                    |

|   |                    |
|---|--------------------|
| <b>Formální a jazyková úroveň, rozsah práce</b>   | <b>A - výborně</b> |
| <i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>   |                    |
| Práce má dobrou logickou strukturu, kvalitní typografii i dobrý anglický jazyk. Diplomantka srozumitelně vysvětluje genetickou motivaci a popisuje svoje řešení. Typografické a jazykové chyby jsou řídké (mj. Fig.2, řádek 3, AT by mělo být |                    |

nahrazeno TT, sec 1.1.1 „are classified are generally“, sec 1.2 „specifying weather“, sec. 5.3.1 „complexity is linearly depend on number the of markers“). Rozsah práce je průměrný, ale v textu nejsou zřejmá žádná významná opomenutí. Délka implementace je několik stovek řádků v jazyce Python, dokumentace kódu je na základní úrovni.

**Výběr zdrojů, korektnost citací**

**B - velmi dobře**

*Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.*

Diplomantka pracuje s relevantními a aktuálními zdroji v přiměřeném rozsahu. Nestandardní a málo informovaný je pouze formát referencí. Většinou chybí standardní bibliografické údaje jako je např. rok publikace nebo číslo.

**Další komentáře a hodnocení**

*Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.*

Vložte komentář (nepovinné hodnocení).

**III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE**

*Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.*

U rovnice 4.9 diplomantka uvádí, že vztah pro uvedenou podmíněnou pravděpodobnost vychází z rovnice 4.1. Tento vztah mi není zřejmý, rovnici 4.9 nepokládám z obecných důvodů za správnou. Můžete prosím podrobněji vysvětlit jak rovnice vznikla?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **A - výborně**.

Datum: 12/06/2018

Podpis: