

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Shlukování RNA-seq reads podle genové exprese
Jméno autora:	Hana Mertanová
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra kybernetiky
Vedoucí práce:	Petr Ryšavý, M.S.
Pracoviště vedoucího práce:	IDA, Katedra počítačů, FEL

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	mimořádně náročné
---------------	--------------------------

Hodnocení náročnosti zadání závěrečné práce.

Studentka pracovala na shlukování sekvenačních dat získaných metodou RNA-seq na základě odhadů exprese. Při tom bylo třeba provést mnoho navazujících kroků – data není možná shlukovat přímo, bylo nutné použít reprezentaci ve formě kratších podřetězců, tzv. k-merů, převést shluky zpět na původní data, dále bylo třeba zprovoznit referenční řešení a navrhnout a porovnat výsledky obou metod. Při tom bylo nutné zpracování velkého množství dat, jde o 62 milionů sekvencí, ze kterých je třeba vybrat malé množství podstatných. Kromě toho byla velká část technik použitých v práci byla pro studentku zcela nová a jsou náplní magisterských předmětů Bioinformatika a Statistická analýza dat. Zadání patří mezi nejtěžších 10 až 20 %, takže hodnotím jako mimořádně náročné.

Splnění zadání	splněno
-----------------------	----------------

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.

Zadání bylo splněno. Body 1-3 jsou pokryty v prvních šesti kapitolách, bod 4 lze najít v kapitole 7, bod 5 odpovídá kapitole 9, bod 6, hlavní těžiště práce, je popsán v kapitolách 8 a 11, bod 7 v kapitolách 10 a 12 a bod 8 lze nalézt v kapitole 13.

Aktivita a samostatnost při zpracování práce	A - výborně
---	--------------------

Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.

Zde nemám nejmenších výhrad. Studentka byla během řešení aktivní, na konzultace chodila vždy pečlivě připravená a mé připomínky průběžně zapracovávala. Bakalářská práce i řešerše během semestrálního projektu vznikaly průběžně během celého semestru. Během semestru nebyly třeba žádné průběžné termíny, studentka se o konzultace hlásila sama vždy po vzniku části práce.

Odborná úroveň	A - výborně
-----------------------	--------------------

Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.

Práci považuji za zdařilou po odborné úrovni. Práce neobsahuje faktografické chyby, a to ani přes to, že studentka pro tvorbu práce musela nastudovat velké množství literatury, která nezahrnuje pouze informatiku, ale i molekulární biologii, buněčnou biologii a genetiku.

Formální a jazyková úroveň, rozsah práce	A - výborně
---	--------------------

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.

V této části bych měl drobné výhrady k použité notaci v některých částech, především co se týká výpočtu chyby. Studentka se snažila být konzistentní s kódem, díky čemuž je notace méně přehledná, než by mohla být. Z hlediska typografické

stránky lze nalézt v práci jen velmi malé množství překlepů, připomínky, které jsem měl byly již zapracovány. Práce je psaná dobrou angličtinou a během tvorby práce jsem pozoroval velký kvalitativní posun v úrovni textu.

Výběr zdrojů, korektnost citací

A - výborně

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.

Práce obsahuje velké množství citací a odkazů na relevantní literaturu. Studentka správně cituje nejenom relevantní články a další literaturu, ale i použité knihovny. Převzaté obrázky jsou odcitovány se zdrojem, vlastní myšlenky jsou dostatečně oddělené od částí, které představují rešerši.

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.

Hodnota chyby E_3 naznačuje, že velké množství sekvencí ve shlucích chybí. Podobně hodnota chyby E_2 nevychází příliš signifikantní. Metoda by zasluhovala ještě nějaká vylepšení, jak studentka naznačuje ve kapitole „Future Work“. Pokud by se podařilo metodu vylepšit, myslím, že by byla vhodná k publikaci. Na základě vlastních experimentů na datech poskytnutých po odevzdání práce studentkou očekávám výraznější problém spíše ve výpočtu chyby (popř. jejím návrhu) než ve vlastní metodě.

Implementační část vhodně kombinuje vlastní kód v Pythonu s knihovnamy pro učení bez učitele a zpracování genomických sekvencí. Po drobných úpravách cest a instalaci potřebného softwaru je možné pustit kódy studentky na metacentru pomocí přiložených skriptů.

III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.

Práci považuji za obhájitelnou nejenom jako bakalářskou práci, ale i jako diplomovou práci. Studentka v práci vhodně kombinuje nastudované znalosti ze strojového učení a bioinformatiky. Zadání považuji za náročné nejenom rozsahem znalostí, které musela studentka získat, ale i objemem zpracovávaných dat a nutností zkombinovat velké množství knihoven a programů. Práce je psaná srozumitelně a dobrou angličtinou a splňuje veškeré formální požadavky i typografické a citační standardy.

Především bych rád vyzdvihнул přístup studentky při tvorbě bakalářské práce. Téma vybírala v předstihu, ještě o prázdninách. Obsah vznikal průběžně a vše bylo hotové několik dní před termínem odevzdání. O konzultace se studentka hlásila sama a chodila na ně připravená.

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

Datum: 30.5.2019

Podpis: