

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Efektivní výpočet podobnosti sekundárních struktur RNA
Jméno autora:	Bc. Marek Hrvol
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	RNDr. Martin Pospíšek, Ph.D.
Pracoviště oponenta práce:	Přírodovědecká fakulta Univerzity Karlovy, Katedra genetiky a mikrobiologie

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	mimořádně náročné
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Cílem práce bylo vytvoření heuristického nástroje pro porovnávání sekundárních struktur ribonukleových kyselin (RNA). Zadání je z mého pohledu náročné. Student byl postaven před úkol využít mělkých neuronových sítí pro porovnání dvou a více sekundárních struktur najednou, případně nalezení podobných sekundárních struktur v rozsáhlých souborech sekundárních struktur RNA. Kromě technických dovedností byl student pro úspěšné zvládnutí projektu nucen se do určité míry seznámit též se zákonitostmi stavby nukleových kyselin, zejména RNA, a získat detailní přehled o nástrojích v současnosti používaných pro porovnávání sekundárních struktur RNA.	

Splnění zadání	splněno
<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>	
Zadání student splnil. Podle mého soudu došlo dokonce k rozšíření zadání o implementaci použitých metod a vzniku aplikace v programovacím jazyku Python, která umožňuje neoborníkům použití algoritmu Struct2Vec pro porovnávání, respektive vyhledávání, sekundárních struktur RNA.	

Zvolený postup řešení	vynikající
<i>Posuďte, zda student zvolil správný postup nebo metody řešení.</i>	
Zvolený postup řešení je originální a nekopíruje žádný ze známých a zavedených postupů. Velmi oceňuji způsob převedení sekundární struktury RNA na slova, což následně umožnilo využití algoritmů používaných pro práci s texty.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Práce je na vysoké odborné úrovni. Zejména oceňuji propojení znalostí algoritmů a programování se znalostmi v oblasti struktury nukleových kyselin, což nebývá obvyklé.	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Grafická úprava práce je dobrá. Ze zvýšeného množství různých překlepů je vidět určitý spěch při dokončování práce. Text je nicméně čtivý a jazykové chyby a překlepy nebrání jeho porozumění.	

Výběr zdrojů, korektnost citací	A - výborně
<i>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</i>	

zvyklostmi a normami.

Student použil relevantní zdroje, které jsou i řádně citovány. Některé způsoby citace použitých pramenů jsou pro přírodovědce nezvyklé. To nicméně nebrání jejich nalezení. Těž bych doporučil číslovat citace postupně podle jejich prvního uvedení v textu.

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.

Řešení projektu i jeho prezentaci považuji za velmi zdařilé. Další komentáře uvádím v celkovém hodnocení níže.

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.

Jak již jsem se zmínil. Práci považuji za velmi povedenou. Rozhodně oceňuji studentův pohyb na rozhraní dvou velmi nepříbuzných oborů a výsledky, které v průběhu řešení projektu získal. V souladu s autorovým návrhem uvedeným v samém závěru práce doporučuji natrénovat a otestovat model na různorodějším souboru dat než poskytují databáze The RNase P database a Comparative RNA Web Site and Project (CRW). Po úspěšném testování na širším okruhu RNA struktur by mohl být navržený postup publikovatelný v renomovaném vědeckém časopise. Z těchto důvodů si dovoluji doporučit pozdržení zveřejnění práce.

K práci mám následující dotazy:

- Návrh transformace sekundární struktury RNA na slova je jednou z klíčových částí práce. Mohl by autor ukázat, zda-li a případně jak je jeho metoda využitelná pro popis pseudouzlu? Domnívám se, že v popisu druhé báze na straně 41 je chyba. Je to drobnost, ale může autor toto komentovat?
- Lze nějak odhadnout a vyčíslit pravděpodobnost s jakou Struct2Vec vrátí shodný výsledek při opakovaném prohledávání (porovnávání) stejné databáze například RNase P db stejnou neznámou strukturou nebo strukturami?
- Je možné v konkrétních nebo i obecných případech porovnání struktur odhadnout, respektive racionálně vysvětlit případný pozitivní vliv redukce počtu slov ("binning") na přesnost výsledků nebo se jedná o empirické zjištění (viz. kapitola 8.3).
- Pro další práci bych doporučil využití databáze RNA struktur Rfam (<https://rfam.xfam.org/>). Mohl by autor komentovat případnou využitelnost dat z této databáze v prostředí Struct2Vec?

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

Datum: 12.6.2019

Podpis: