

POSUDEK VEDOUCÍHO DIPLOMOVÉ PRÁCE

Autor práce: Bc. Marek Hrvol
Název práce: Efektivní výpočet podobnosti sekundárních struktur RNA
Vedoucí: Doc. Ing. Jiří Kléma, Ph.D.

Ribonukleová kyselina (RNA) hraje klíčovou roli v syntéze bílkovin. Dlouhodobě je známo, že je využívána jako pasivní nosič informace mezi DNA a bílkovinami. Nověji je studována i její aktivně regulační funkční role v buňce. U každého biopolymeru je funkce určena především jeho strukturou, tím jak se zavine a jaké zaujme prostorové uspořádání. Sekundární struktura RNA toto zavnutí vyjadřuje zjednodušeně na úrovni spojení jednotlivých bazických párů a lze jí využít k posouzení homologie mezi různými molekulami RNA. Stávající přístupy k porovnání sekundárních struktur RNA jsou relativně pomalé a v mnoha úlohách nepoužitelné. Cílem diplomové práce je tedy navrhnout novou efektivní metodu výpočtu podobnosti sekundárních struktur RNA.

Metoda je založena na přístupu `struc2vec`, který se inspiruje v dobře známém přístupu `doc2vec` používaném pro vytváření numerických reprezentací textových dokumentů. I zde autor nejprve rozkládá sekundární strukturu RNA na slova. Reprezentace struktury pomocí slov je posléze pomocí neuronových sítí převedena na reprezentaci numerickou, sekundární struktury RNA jsou zanořeny do vysokodimenzionálního vektorového prostoru. V něm je výpočet podobnosti mezi sekundárními strukturami trivální. Cílem práce bylo upřesnit výše popsany obecný postup, implementovat jej, srovnat jej s postupy stávajícími z pohledu efektivity a hodnocení podobnosti struktur. Na závěr bylo úkolem diplomanta použít metodu `struc2vec` v modelové úloze vyhledání virových strukturálních motivů v lidském genomu.

Diplomant řešil své téma dlouhodobě. Pravidelně jsme se scházeli přibližně po dobu jednoho akademického roku, za tu dobu nenastaly v řešení žádné větší prodlevy. Rychlost postupu prací byla ve srovnání s jinými diplomanty standardní. Pan kolega Hrvol postupoval samostatně a prokázal dostatečné schopnosti k tomu, aby dosáhl všech cílů určených zadáním. Cíle jsou nicméně splněny přesně na úrovni zadání, diplomant neuvažoval o jejich aktivním upřesňování či prohlubování. Text práce popisuje průběh prací i výsledky dostatečně, obsahuje ale některé nepřesnosti, věcné nedostatky i poměrně velký počet překlepů a jiných jazykových chyb. Ani po odevzdání práce není zcela zřejmé nakolik jsou odchylky mezi `struc2vec` a editační vzdáleností ve prospěch či neprospěch strukturně homologických párů sekvencí. Autor totiž částečně vynuceně zaměňuje pojem podobnosti sekundární struktury za shodu v hrubém taxonomickém zařazení ribonukleových kyselin. Názorné příklady porovnávaných struktur v příloze B naznačují, že relativně nízká shoda s editační vzdáleností nemusí být překážkou širšího využití metody `struc2vec`.

Souhrnně lze říci, že práce splňuje všechny cíle, které jsme si na počátku vytkli. Byl vytvořen a otestován konkrétní efektivní algoritmus pro rychlé porovnávání sekundárních struktur RNA. Práci doporučuji k obhajobě a hodnotím ji známkou

B — velmi dobře.

V Praze, dne 13. června 2019

Doc. Ing. Jiří Kléma, Ph.D., vedoucí