

OPONENTSKÝ POSUDOK NA DIZERTAČNÚ PRÁCU ING. IVANA VOGELA "SINGLE-CELL GENOTYPING"

Predložená dizertačná práca Ing. Ivana Vogela sa zaoberá vývojom nových algoritmov na genotypovanie pomocou génových čipov (SNP arrays) v kontexte jednotlivých buniek. Artefakty laboratórnych postupov v prípade analýzy individuálnych buniek zavádzajú do výsledku väčšie množstvo šumu, ako je obvyklé pri bežnom použití tejto technológie, čo si vyžaduje vývoj nových informatických metód. Jedná sa o interdisciplinárnu prácu, autor sa však v rámci predloženej dizertačnej práce venuje najmä jej informatickým aspektom, takže **práca je pevne zakotvená v odbore výpočtová technika a informatika.**

Hlavným prínosom práce je vývoj nového dvojstupňového algoritmu, ktorý kombinuje neparametrické a parametrické modely strojového učenia. O tomto algoritme autor preukazuje, že umožňuje efektívne odfiltrovať spomínané zvýšené množstvo šumu. Okrem vývoja samotného algoritmu bolo v tomto prípade náročné navrhnuť experimentálne overenie výkonnosti daného riešenia, keďže v tejto oblasti je veľmi ťažké získať správne odpovede (ground truth). Táto časť práce taktiež vykazuje originálne prvky a autor preukazuje, že jeho algoritmus dokáže prekonať doteraz používané riešenia. Autor tiež ukazuje niekoľko nových originálnych aplikácií biologických analýz, ktoré je možné realizovať vďaka lepšej presnosti jeho algoritmu (aj keď v týchto prípadoch zväčša chýba porovnanie, ktoré by demonštrovalo, že doteraz existujúce nástroje na tieto aplikácie nestačia). Celkovo prácu považujem **za originálny a prínosný príspevok ku súčasnému stavu riešenej problematiky.**

Jednou zo slabín predloženej dizertačnej práce je jej samotný text. Práca sa ťažko číta nielen všeobecným odborníkom na informatiku, ale aj čitateľom vzdelaným v oblasti bioinformatiky. Úvodné kapitoly sú príliš všeobecné, aby boli nápomocné pri čítaní práce, pokiaľ čitateľ nie je už odborníkom v danej oblasti. Kapitoly 5 a 6 popisujúce jadro prínosu práce sú čitateľné pre odborníka, no očakával by som väčšiu úroveň detailu a zdôvodnenie rozhodnutí, ktoré autor urobil pri návrhu vlastných riešení. V kapitolách 7 a 8 popisujúcich ďalšie aplikácie autor už voľne používa biologickú terminológiu špecifickú pre danú oblasť bez akéhokolvek úvodu, tieto kapitoly sú pre väčšinu informatikov či bioinformatikov nečitateľné.

Obzvlášť mi v práci chýbala **podrobnejšia charakterizácia chýb**, ktoré vznikajú pri analýze jednotlivých buniek (v porovnaní s klasickým genotypovaním) a hĺbková analýza **prečo sa aktuálne používané metódy nedokážu s takýmito chybami vyrovnáť**. Autor rieši daný problém kombináciou dvoch modelov strojového učenia, ktorých samostatná aplikácia nie je výrazne lepšia ako predchádzajúce riešenia, no ich kombinácia dokáže predchádzajúce riešenia prekonať. **Aká bola motivácia pre aplikáciu dvojstupňového riešenia a prečo dvojstupňové riešenie funguje lepšie?** Bol by som vďačný, ak by sa uchádzač mohol k týmto otázkam vyjadriť v rámci obhajoby.

Jadro dizertačnej práce bolo akceptované ako výskumný článok v uznávanom časopise *Bioinformatics* (epub ahead of print, prijatý v máji 2019). Autor v materiáloch uvádza svoj podiel na tomto článku 90%. Som presvedčený, že tento podiel je adekvátny, čo sa týka analýzy dát, no článok tiež zahŕňa významné množstvo nových experimentálnych dát a z tohto hľadiska mám pochybnosti o tom, že príspevok ostatných autorov ku článku ako celku je limitovaný 10% autorským podielom. Ďalším publikovaným príspevkom priamo prepojeným s dizertačnou prácou je článok v časopise *Science*, kde autor preukázal použitie výsledkov dizertačnej práce na získanie biologických záverov, ktoré tvoria významnú súčasť článku. **Jadro dizertačnej práce bolo publikované adekvátnym spôsobom na potrebnej úrovni.**

Na záver, aj napriek výhradám ku samotnému textu dizertačnej práce konštatujem, že autor v práci a celkovo svojou doterajšou činnosťou preukázal schopnosť samostatnej vedeckej práce v odbore výpočtová technika a informatika. Práca podľa môjho názoru zodpovedá všeobecne uznávaným požiadavkám na udelenie titulu PhD v tomto odbore.

V Bratislave, 15.11.2019

doc. Mgr. Tomáš Vinař, PhD.