

Posudek diplomové práce

Autor práce Nicole Aemilia Urban
Název práce Interactive clustering approaches in single-cell cytometry
Rok odevzdání 2024
Studijní program Bioinformatika

Autor posudku Adam Šmelko **Role** vedoucí
Pracoviště D3S MFF UK

Text posudku:

Analýza cytometrických dát je důležitým nástrojem pro studium biologických systémů a v této doméne bylo publikovaných mnoho komplexních algoritmov, jako hierarchický clustering s Mahalanobisovou vzdialenosťou. Táto práca sa snaží priblížiť moderné metódy analýzy cytometrických dát užívateľom bez programátorských skúseností. Existuje mnoho nástrojov na vizualizáciu cytometrických dát, ale netriviálna časť z nich je closed-source, vyžadujú mieru znalosti programovacích jazykov, alebo nedisponuje bežnými interaktívnymi prvkami. Praktická časť tejto práce sa zameriava práve na tieto problémy s cieľom zjednodušiť doménovým expertom analýzu cytometrických dát. Hlavný prínos práce je vizualizačný nástroj Ash, ktorý umožňuje interaktívnu analýzu dát z hierarchického clusteringu.

V prvej kapitole študentka vysvetľuje prietokovú cytometriu, históriu analýzy cytometrických dát, clustering a spôsoby vizualizácie.

Hlavná časť práce sa venuje návrhu a implementácii Ash. Druhá kapitola začína rozborom súvisiacej práce a snaží sa vymedziť Ash od existujúcich vizualizačných nástrojov. Asi najväčší rozdiel je, že Ash neprijíma ako vstup FCS súbor, ale dendrogram, čo aplikáciu nelimituje len na clustering prítomný v danom vizualizačnom nástroji. Ďalej sa študentka venuje dôkladnej analýze vybraných softwarových technológií a diskutuje ich výhody a nevýhody. V závere kapitoly študentka popisuje implementáciu a užívateľské rozhranie Ash. Zaujímavý je algoritmus na 'splitovanie' dendrogramu, inšpirovaný iDendro a ShinySOM, ktorý ku známym spôsobom vizualizácie (ako heatmap, scatterplot, ...) pridáva nový interaktívny prvok.

Posledná kapitola aplikuje Ash na reálnom datasete. Študentka pomocou splitovania dendrogramu prepojeného s UMAP postupne identifikuje chybu v manuálnom gatingu.

Hlavným prínosom práce je software, ktorý spája najpoužívanejšie vizualizačné nástroje spoločne s navrhnutými interaktívnymi prvkami. Študentka dbá na jednoduchosť nasadenia a ako príklad poskytuje hostovanú verziu Ash dostupnú cez webový prehliadač. Aplikovaním splitovania dendrogramu na reálnych dátach študentka demonštruje užitočnosť Ash ako nástroja pre interaktívnu

analýzu cytometrických dát.

V práci sa objavuje niekoľko faktických chýb alebo tvrdení, ktorých význam je ťažké odhadnúť. Napr. na strane 12 pri prirovnávaní Mahalanobisovej vzdialenosti 'variance of each equal zero' (asi to malo byť myslené variance is unit alebo mean is zero). Za ďalší nedostatok považujem nedostatočné vymedzenie sa voči existujúcim nástrojom v kapitole 2.1. Síce tomu študentka venuje jeden paragraf, myslím si, že pre neznaleho čitateľa by bolo užitočné dôkladnejšie vysvetlenie. Rušivým prvkom sú gramatické chyby, nedostatočná formálna úprava (strana 45) alebo občasná nejednoznačnosť v texte (poznámky pod čiarou na strane 29 a 30), ktoré by sa dali odstrániť dôkladnejšou korektúrou textu.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 16. 8. 2024

Podpis: