

Cvičenia pre biológov, 7.12.2023

Zhrnutie semestra

Tvorba bioinformatického nástroja

- Sformulujeme biologické ciele
(aké máme dáta, aké typy otázok sa chceme pýtať).
- Sformulujeme informaticky/matematicky
(napr. ako pravdepodobnostný model).
Dostaneme informatické zadanie problému, v ktorom je presne daný vzťah medzi vstupom a želaným výstupom
(napr. nájsť zarovnanie s max. skóre v určitej skórovacej schéme).
- Hľadáme efektívne algoritmy na riešenie informatického problému.
- Ak sa nám nepodarí nájsť dosť rýchly algoritmus, použijeme heuristiky, ktoré dávajú približné riešenia.
- Testujeme na reálnych dátach, či sú výsledky biologicky správne
(či bol model dobre zvolený, či heuristiky dobre fungujú).

Použitie bioinformatického nástroja

- Sformulujeme biologické ciele (aké máme dáta, aké typy otázok sa chceme pýtať).
- Porozmýšľame, aký typ nástroja, resp. ich kombinácia by nám mohli pomôcť
- Alebo hľadáme v literatúre nástroj na typ problému, s ktorým sme sa ešte nestretli
- Pre správne nastavenie parametrov a interpretovanie výsledkov je dôležité poznať model, predpoklady, ktoré autori nástroja použili, resp. zdroj dát v príslušnej databáze
- Konkrétne nástroje a webstránky sa rýchlo menia, celkové princípy sa menia pomalšie

Prehľad preberaných tém

- Zostavovanie genómov (najkratšie spoločné nadslovo, heuristiky, de Bruijnov graf)
- Zarovnanie (skórovanie ako pravdepodobnostný model, dynamické programovanie, heuristické zarovnávanie, E-value a P-value, lokálne vs. globálne, párové vs. viacnásobné, celogenómové)
- Evolúcia (pravdepodobnostné modely substitúcií, metóda maximálnej vierohodnosti, metóda maximálnej úspornosti, metóda spájania susedov)
- Hľadanie génov (skryté Markovove modely)
- Komparatívna genomika (hľadanie konzervovaných oblastí, komparatívne hľadanie génov, pozitívny výber, fylogenetické HMM, kodónové matice)

Prehľad preberaných tém (pokračovanie)

- Expresia génov (zhlukovanie, klasifikácia, regulačné siete, transkripčné faktory, hľadanie motívov)
- Proteíny (predikcia štruktúry, profily a profilové HMM rodín/domén, protein threading)
- RNA štruktúra (dynamické programovanie, stochastické bezkontextové gramatiky)
- Populačná genetika (mapovanie asociácií, väzbová nerovnováha, genetický drift, štruktúra a história populácie)

Nahliadli sme do sveta informatiky

- Algoritmus, časová zložitosť
- NP-ťažké problémy, presné algoritmy, heuristiky, aproximačné algoritmy
- Dynamické programovanie
- Stromy, grafy
- Skryté Markovove modely a bezkontextové gramatiky

Ďalšie predmety

- **Genomika** N-mCBI-303, Nosek a kol. (LS, 2P, 3kr)
- **Linux pre používateľov** 1-AIN-500, Nagy (LS, 2K, 2kr) alebo
Operačné systémy a počítačové siete 1-DAV-103 (ZS, 2P+2C, 5k)
- **Programovanie (1)** 1-MAT-130, Salanci (ZS, 2P+2C, 5kr) alebo
Programovanie (1) 1-AIN-130 Blaho (ZS, 4P+4C, 9kr)

Pre pokročilejších

- **Seminár z bioinformatiky 1, 2** Brejová, Vinař (ZS/LS, 2S, 2kr)
journal club o bioinformatických metódach
- **Úvod do bioštatistiky** 1-BMF-331 Waczulíková (LS, 2P+1C, 4kr)
predmet pre biomedicínskych fyzikov a dátovú vedu
- **Vizualizácia dát** 1-DAV-105 Brejová, Bátorová (LS, 2P+2C, 5kr)
vyžaduje základy Pythonu
- **Manažment dát** 1-DAV-105 Brejová, Boža, Vinař (LS, 1P+2C, 5kr)
vyžaduje znalosť programovania, základy práce na príkazovom riadku