

Detekcja interakcji w dużych zbiorach potencjalnych predyktorów

Jan Mielniczuk
Instytut Podstaw Informatyki PAN, Warszawa
miel@ipipan.waw.pl

1. Opis projektu

Celem projektu jest zaproponowanie nowych efektywnych metod selekcji istotnych interakcji epistatycznych w danych w Genome-Wide Association Study (GWAS) i porównanie ich z metodami obecnie stosowanymi. Dane GWAS dotyczą związków między genotypami i fenotypami w grupach będących obiektem badania. Podstawowym problemem będzie wykrywanie interakcji występującej między dwoma Single Nucleotide Polymorphisms (SNPs) w określeniu wartości związanej cechy binarnej. Trzy, częściowo nierozwiązane zagadnienia, które są przedmiotem projektu, to: kontekstowość określenia interakcji, metody jej pomiaru oraz wykonalne obliczeniowo algorytmy pozwalające uniknąć sprawdzania wszystkich par SNP. Badania dotyczące pierwszego i drugiego problemu są już prowadzone i skupiają się na nieparametrycznej mierze Informacji Interakcyjnej, która pozwala często wykrywać interakcje pomijane przez metody bazujące na modelach parametrycznych. Istotne nowe problemy tutaj to między innymi problem określenia jej przybliżonego rozkładu w przypadku występowania jedynie efektów głównych. Trzeci problem ma zasadnicze znaczenie i jest obiektem intensywnych badań na świecie. Poszukiwana metoda dwustopniowa składająca się z etapu odsiewania i selekcji powinna pozwolić zachować po odsianiu pary genów bez efektów głównych, ale z silną interakcją. Skonstruowany algorytm powinien wykazywać istotną przewagę nad obecnie stosowanymi metodami takim jak m.in. TEAM, BOOST, SNPHarvester and Screen and Clean. Praca Upton et al, Review: High-performance computing to detect epistasis in genome scale data sets, Briefings in Bioinformatics 2015, zawiera przegląd tematyki i referencje.

2. Wymagania

- a. Ukończone studia drugiego stopnia z informatyki, matematyki lub fizyki
- b. Umiejętności programistyczne
- c. Podstawowa wiedza dotycząca uczenia maszynowego, statystyki, metod numerycznych i optymalizacji
- d. Ciekawość poznawcza i chęć rozwiązywania problemów
- e. Dobra znajomość języka angielskiego