

ANALIZA BIOINFORMATYCZNA MEDYCZNYCH BAZ DANYCH

prof. dr hab. Jacek Waniewski, dr inż. Małgorzata Dębowska

Instytut Biocybernetyki i Inżynierii Biomedycznej PAN

Zakład IV

Pracownia Modelowania Matematycznego Procesów Fizjologicznych

Obecnie na świecie notuje się bardzo szybki przyrost danych medycznych. Pojedynczy pacjent czy grupa pacjentów opisywana jest za pomocą zbioru kilkudziesięciu, kilkuset, czy w przypadku danych genetycznych, kilku tysięcy/milionów cech. Modele statystyczne najczęściej dedykowane są do sytuacji, kiedy liczba pacjentów znacznie przewyższa liczbę cech, aby uzyskać statystyczną istotność. Ponadto w praktyce często ocenia się znaczenie pojedynczych cech, które to podejście ignoruje możliwość interakcji cech podkreślane przez klinicystów. Celem planowanej pracy doktorskiej jest zastosowanie różnorodnych metod bioinformatycznych i statystycznych do analizy dużych i złożonych zbiorów danych medycznych. W szczególności analizowane będą dane antropometryczne, demograficzne, biomarkery oznaczane we krwi, wyniki interwencji terapeutycznych, dane genetyczne pacjentów z różnymi schorzeniami. W uzupełnieniu do typowej analizy statystycznej planowane jest zastosowanie metod bioinformatycznych, jak Lasso (ang. *least absolute shrinkage and selection operator*) czy maszyna wektorów nośnych (ang. *support vector machine*, SVM). Użycie powyższych metod pozwala na wybór zbioru cech związanych ze specyficznym problemem klinicznym oraz na przeprowadzenie klasyfikacji. Zastosowanie odpowiedniej metody selekcji cech powinno pozwolić na zredukowanie liczby cech poprzez pominięcie cech nieistotnych lub redundantnych dla rozważanego problemu. Zakłada się badanie powiązań i interakcji między cechami. Dobór odpowiedniej metody analizy do danego zbioru danych jest kluczowy. Dla pewnych zbiorów danych będzie korzystne zastosowanie metody typu 'deep learning', w innym przypadku odpowiednią metodą będzie liniowa regresja. Metodyczne i kompleksowe podejście do obróbki medycznych baz danych, odpowiednia wizualizacja, pozwalają na właściwą interpretację kliniczną. Potencjalne, praktyczne zastosowania uzyskanych informacji zawierają trafną diagnozę oraz optymalizację terapii dla pojedynczego pacjenta. Praca będzie realizowana we współpracy międzynarodowej z Instytutem Karolińskim w Sztokholmie.