

Dynamiczny rozwój współczesnej medycyny, biologii i chemii opiera się na skomplikowanych technikach pomiarowych takich jak spektrometria mas i spektroskopia magnetycznego rezonansu jądrowego. Obydwie techniki umożliwiają poznanie składu złożonych substancji, takich jak leki, płyny ustrojowe, tkanki ludzkie lub sztucznie stworzone polimery.

Choć zjawiska fizyczne jakie wykorzystują te techniki są różne, to okazuje się, że uzyskane dane pomiarowe można analizować za pomocą podobnych metod. Opracowanie algorytmów do analizy danych spektrometrycznych jest niezwykle ważne, gdyż są to techniki wysoko-przepustowe, co oznacza, że mamy do czynienia z gigabajtami danych dla pojedynczego eksperymentu, czyli nie jest możliwa, inna niż automatyczna, analiza i interpretacja widm.

Jako przykład rozważmy obrazowanie spektrometryczne, w którym badamy skład przestrzenny tkanki, czyli każdy piksel zawiera informację o tysiącach metabolitów i peptydów znajdujących się w danym miejscu. Innym przykładem jest widmo magnetycznego rezonansu jądrowego dla określonego leku, który zawiera kilka lub kilkanaście substancji (zarówno substancje czynne, jak też substancje pomocnicze). Bardzo złożone widma uzyskujemy dla syntetycznych polimerów, które zawierają łańcuchy o różnych długościach i różnych grupach bocznych.

W wymienionych przykładach głównym zadaniem dla informatyka jest zidentyfikować substancje, które znajdują się w próbce oraz oszacować ich proporcje. W pierwszym przypadku nie wiemy jakie molekuly znajdują się w tkance, więc musimy posłużyć się całym zestawem potencjalnych substancji. W sytuacji leku dysponujemy wiedzą o jego składzie, jednak proporcje substancji mogą być zmienione i analiza widma powinna to wykryć, podobnie jak ewentualne zanieczyszczenia preparatu.

Założmy, że widmo badanej substancji jest ciągiem słupków ułożonych na osi liczb dodatnich. W przypadku widma masowego położenie słupka odpowiada masie molekularnej substancji, a wysokość (traktowana jako masa słupka) intensywności sygnału. Nasze algorytmy dla porównania dwóch takich widm używają pojęcia optymalnego transportu, które można uznać za uogólnienie problemu sortowania. Optymalny transport to scenariusz, który przesuwa słupki z jednego widma, tak aby uzyskać drugie minimalizując wykonaną pracę, czyli iloczyn przesuwanej masy i drogi jaką trzeba pokonać.

Optymalny transport pozwala na znalezienie odległości między widmami. Okazuje się, że taka odległość ma wiele dobrych własności i adekwatnie porównuje widma. Używając tej odległości potrafimy rozwiązać problem dekompozycji złożonego widma na pojedyncze składowe substancje za pomocą regresji liniowej. Taki problem zamierzamy rozwiązać dla tysięcy metabolitów w obrazowaniu spektrometrycznym, dla złożonych widm leków oraz dla trudnych w analizie, ze względu na nachodzące na siebie sygnały, widm syntetycznych polimerów.