

Modele filogenetyczne do analizy ewolucji komórek rakowych

Postępy w technologiach eksperymentalnych umożliwiają obecnie analizę olbrzymiej ilości danych uzyskiwanych z sekwencjonowania materiału genetycznego pojedynczych komórek. Komórki w guzie nowotworowym można podzielić na klony, czyli populacje dzielące wspólne cechy, które są wynikiem mutacji. Badania nad różnorodnością komórek rakowych miały miejsce już w XIX wieku, kiedy to pod mikroskopem zaobserwowano morfologiczne różnice pomiędzy pojedynczymi komórkami z jednego guza nowotworowego. W dzisiejszych czasach, mamy możliwości wykorzystywać takie badania do podawania prognozy, wyjaśniania mechanizmów lekooporności, oraz dopasowywania terapii.

Metodą obrazowania ewolucji komórek nowotworowych u pojedynczego pacjenta jest drzewo filogenetyczne, w którym liście reprezentują pojedyncze komórki, a węzły wewnętrzne symbolizują ich hipotetycznych wspólnych przodków. Pojedyncza komórka może być reprezentowana przez profil liczności kopii, ang. copy-number profile (CNP), czyli liczności kopii fragmentów chromosomu referencyjnego. W naszym projekcie zakładamy, że profile CNP są dane.

W celu dokonywania porównań materiału genetycznego dwóch komórek niezbędna jest miara, która określa jak bardzo dwa profile CNP różnią się od siebie. Jednym z podejść jest dystans ewolucyjny, czyli liczba zdarzeń ewolucyjnych niezbędna do przekształcenia jednego profilu CNP w drugi. Nie jest to odległość w matematycznym rozumieniu, gdyż nie jest symetryczna. W procesie ewolucyjnym komórek nowotworowych duże części genomu mogą zostać zwielokrotnione lub usunięte. Najnowsza z zaproponowanych odległość ewolucyjna, opublikowana w 2020, modeluje to zjawisko. Jednakże, obliczenie jej wartości to złożony problem, który w informatyce określa się jako NP-trudny. Wcześniejsze uproszczone warianty dystansów ewolucyjnych są prostsze do obliczenia, ale problem rekonstrukcji drzew filogenetycznych, nawet dla tych uproszczonych modeli, jest złożony (NP-trudny). Podsumowując, mierzymy się ze złożonym problemem, który składa się ze złożonych podproblemów, gdy tymczasem jest potrzeba obliczeniowo wykonalnych modeli i rozwiązań.

Nasz projekt skoncentruje się na zaprojektowaniu i analizie:

- Modeli filogenetycznych dla danych uzyskanych z regularnych badań pojedynczego pacjenta,
- Modeli wyszukujących prawidłowości w procesach ewolucyjnych dla wielu drzew filogenetycznych.

Pierwszy z powyższych tematów jest szczególnie innowacyjny, gdyż drzewa filogenetyczne modelują przeszłość z danych dostępnych w teraźniejszości. Przykładowo, do modelowania ewolucji możemy analizować genomy tylko obecnie występujących gatunków, a dane paleontologiczne nie zapewnią nam kompleksowego obrazu przeszłości. Dane pobrane od pojedynczego pacjenta w regularnych badaniach zapewnią informację o wszystkich populacjach komórek istniejących w czasie przeprowadzanego badania. Rodzi się przed nami nowa możliwość, aby modelować ewolucję pomiędzy punktami kontrolnymi. Procesy ewolucyjne zachodzą tak szybko, że możemy je zaobserwować.

Nasza analiza teoretyczna będzie wykorzystywać standardowe techniki z zakresu dowodzenia matematycznego, teorii grafów, genomiki porównawczej, rekonstrukcji drzew filogenetycznych. Dla wszystkich modeli zaprojektujemy i zaimplementujemy algorytmy. Nasze badania będą wykorzystywać standardowe algorytmiczne paradygmaty projektowania, analizy złożoności działania i złożoności problemu. Spodziewamy się napotkania problemów NP-trudnych, dla których rozwiązania wykorzystamy heurystyki, algorytmy optymalizacyjne i aproksymacyjne.