Algorytmy kombinatoryczne w bioinformatyce, wykładowca prof. Marta Kasprzak

Materiały uzupełniające do wykładu 1: wprowadzenie.

SLAJD 2

Treści przekazywane w ramach wykładów mieszczą się w obrębie tematyki wyszczególnionej hasłowo na tym slajdzie. Wykład 1 ma charakter wprowadzający, w dużej mierze odwołuje się do pojęć już znanych, m.in. z przedmiotów Matematyka dyskretna i Algorytmy i struktury danych.

SLAJD 5

Chociaż pojęcia „bioinformatyka" i „biologia obliczeniowa" bywają stosowane zamiennie, można zaobserwować różnicę w ich znaczeniu, zwłaszcza w ujęciu historycznym. Oba pojęcia mogą pokrywać rozwiązania z zakresu informatyki zarówno bardziej teoretycznej (metody, modele matematyczne, twierdzenia), jak i praktycznej (programy, narzędzia, bazy danych), szala znaczenia przeważała jednak na stronę teoretyczną w przypadku biologii obliczeniowej i praktyczną w przypadku bioinformatyki. Niejakim potwierdzeniem takiego przyporządkowania znaczenia była niegdyś zawartość publikacji w dwóch najbardziej liczących się czasopismach z tego obszaru, *Bioinformatics* i *Journal of Computational Biology*. Obecnie różnica pomiędzy tymi pojęciami zaciera się, podobnie treści publikowane w wwym. czasopismach. Wykłady z tego przedmiotu koncentrują się na wybranych algorytmach i modelach matematycznych opracowanych dla problemów biologicznych.

SLAJD 6

Wymienione w ostatnim podpunkcie problemy omawiane są w ramach innych przedmiotów. Problem komiwojażera przedstawiony jest na slajdach 18–19 bieżącego wykładu, formalny zapis problemu plecakowego można znaleźć na slajdzie 23 wykładu zamieszczonego tutaj:
http://www.cs.put.poznan.pl/mkasprzak/zp/ZP-wyklad5.pdf

SLAJD 13

Jeśli relację znajomości pomiędzy użytkownikami serwisów społecznościowych przedstawimy w postaci sieci powiązań, tzn. grafu nieskierowanego, gdzie krawędziami zostaną połączone pary wierzchołków odpowiadające parom znajomych, kliki w takim grafie reprezentować będą podgrupy osób znających się każda z każdą. Wyszukiwanie takich podgrup, zwłaszcza dużych, może być celowe z punktu widzenia marketingu. Rozpoznawanie wzorców w obrazie odbywa się zwykle z użyciem innych podejść; w tym ujęciu grafowym porównywany jest graf nieskierowany odpowiadający wzorcowi z większym grafem nieskierowanym odpowiadającym obrazowi. W większym grafie wyszukiwany jest podgraf izomorficzny do grafu wzorca, tzn. o identycznej strukturze dla pewnego przyporządkowania wierzchołków podgrafu i grafu wzorca. W trzecim przykładzie nieskierowany graf oddaje sieć połączeń między wsiami, gdzie krawędziom przypisana jest waga odpowiadająca odległości pomiędzy daną parą wsi. Minimalne drzewo rozpinające jest takim spośród wszystkich możliwych drzew rozpinających grafu, którego suma wag krawędzi składowych jest najmniejsza. Tutaj drzewo takie odpowiadać będzie sieci energetycznej o najkrótszej sumarycznej długości kabli obejmującej wszystkie wsie. Jest to uproszczone spojrzenie na ten problem, w rzeczywistości na postać sieci energetycznej mają wpływ także inne uwarunkowania.

SLAJD 18

W innym przykładowym problemie optymalizacyjnym, poszukiwaniu minimalnego drzewa rozpinającego, funkcją celu jest suma wag krawędzi wchodzących w skład drzewa i jest ona minimalizowana. Rozwiązaniem dopuszczalnym jest każde drzewo rozpinające grafu będącego instancją tego problemu, niezależnie od wag jego krawędzi. Rozwiązaniem optymalnym jest takie drzewo rozpinające tego grafu, które spośród wszystkich innych jego drzew rozpinających ma najmniejszą wartość funkcji celu.

SLAJD 19

Cykl Hamiltona to cykl zawierający każdy wierzchołek grafu dokładnie raz.

SLAJD 27

Algorytm heurystyczny może mieć złożoność wykładniczą, może też mieć złożoność wielomianową, lecz mimo to liczyć zbyt długo (ze względu na wysoki stopień wielomianu lub duże stałe współczynniki). Skoro jednak algorytm taki ma być działającą sprawnie alternatywą dla algorytmu dokładnego, powinien być konstruowany z dbałością zarówno o jakość zwracanych rozwiązań, jak i o czas obliczeń. Wypośrodkowanie pomiędzy tymi dwoma sprzecznymi kryteriami (maksymalizacja jakości, minimalizacja czasu) nie jest zadaniem łatwym. Proste heurystyki działają zwykle szybko, lecz jakość generowanych przez nie rozwiązań często nie jest satysfakcjonująca. Im bardziej złożone heurystyki, tym lepszych rozwiązań można oczekiwać, za to czas obliczeń wzrasta.

SLAJD 28

Wystarczy, że w problemie komiwojażera damy na wejściu macierz niekompletną, gdzie pomiędzy niektórymi miastami nie będzie możliwe bezpośrednie przejście, żeby wygenerowanie rozwiązania dopuszczalnego stało się obliczeniowo trudne. Macierz taka przekłada się na graf niebędący pełnym, w którym poszukiwany jest cykl Hamiltona o minimalnym koszcie. Rozwiązaniem dopuszczalnym jest każdy cykl Hamiltona w takim grafie, ale wiemy, że znalezienie choćby jednego jest problemem NP-trudnym. Heurystyka dla tego problemu może więc nie zwrócić żadnego rozwiązania, bo może nie znaleźć w ograniczonym czasie żadnego rozwiązania dopuszczalnego.