W1

Projektowanie efektywnych algorytmów – wykład W04ITE-SM0066G r.a. 2024/2025 ECTS 4

Prowadzący: dr inż. Tomasz Kapłon p. 312 C3 tomasz.kaplon@pwr.edu.pl

Forma zaliczenia: Zaliczenie na ocenę

Spotkania (piątki): 11:15 – 13:00

04/10, 11/10, 18/10, 25/10, 08/11, 22/11, 29/11, 06/12, 11/12, 13/12, 20/12, 10/01, 17/01, 24/01, 03/02

Problem

Dlaczego niektóre problemy są trudne* do rozwiązania?

* nie ciężkie ☺

Problem

Dlaczego niektóre problemy są trudne do rozwiązania?

1. Niektóre problemy są skomplikowane (posiadają wiele zmiennych wpływających na wynik i przebieg rozwiązania), i nawet jeśli znamy metodę ich rozwiązywania, to konieczne jest ich uproszczenie (zmniejszenie liczby argumentów [zmiennych w problemie] bądź zmniejszenie ograniczeń [relaksacja]), co prowadzić może do tego, że uzyskane rozwiązania są bezużyteczne.

Problem

Dlaczego niektóre problemy są trudne do rozwiązania?

1. Niektóre problemy są skomplikowane (posiadają wiele zmiennych wpływających na wynik i przebieg rozwiązania), i nawet jeśli znamy metodę ich rozwiązywania, to konieczne jest ich uproszczenie (zmniejszenie liczby argumentów [zmiennych w problemie] bądź zmniejszenie ograniczeń [relaksacja]), co prowadzić może do tego, że uzyskane rozwiązania są bezużyteczne.
2. Rozmiar przestrzeni rozwiązań (liczba rozwiązań dopuszczalnych) jest tak rozległy, że rozwiązanie poprzez sprawdzenie wszystkich możliwości jest nieefektywne obliczeniowo (czasowo bądź pamięciowo).

Problem

Dlaczego niektóre problemy są trudne do rozwiązania?

1. Niektóre problemy są skomplikowane (posiadają wiele zmiennych wpływających na wynik i przebieg rozwiązania), i nawet jeśli znamy metodę ich rozwiązywania, to konieczne jest ich uproszczenie (zmniejszenie liczby argumentów [zmiennych w problemie] bądź zmniejszenie ograniczeń [relaksacja]), co prowadzić może do tego, że uzyskane rozwiązania są bezużyteczne.
2. Rozmiar przestrzeni rozwiązań (liczba rozwiązań dopuszczalnych) jest tak rozległy, że rozwiązanie poprzez sprawdzenie wszystkich możliwości jest nieefektywne obliczeniowo (czasowo bądź pamięciowo).
3. Liczba rozwiązań dopuszczalnych jest tak mała (ze względu na warunki [ograniczenia] zadania), że trudno jest wygenerować rozwiązanie w ogóle, nie mówiąc o uzyskaniu rozwiązania optymalnego.

Problem

Dlaczego niektóre problemy są trudne do rozwiązania?

1. Niektóre problemy są skomplikowane (posiadają wiele zmiennych wpływających na wynik i przebieg rozwiązania), i nawet jeśli znamy metodę ich rozwiązywania, to konieczne jest ich uproszczenie (zmniejszenie liczby argumentów [zmiennych w problemie] bądź zmniejszenie ograniczeń [relaksacja]), co prowadzić może do tego, że uzyskane rozwiązania są bezużyteczne.
2. Rozmiar przestrzeni rozwiązań (liczba rozwiązań dopuszczalnych) jest tak rozległy, że rozwiązanie poprzez sprawdzenie wszystkich możliwości jest nieefektywne obliczeniowo (czasowo bądź pamięciowo).
3. Liczba rozwiązań dopuszczalnych jest tak mała (ze względu na warunki [ograniczenia] zadania), że trudno jest wygenerować rozwiązanie w ogóle, nie mówiąc o uzyskaniu rozwiązania optymalnego.
4. Sposób oceny rozwiązania jest niepoprawny, nieprecyzyjny (niewłaściwie określona funkcja celu [kosztu, nagrody, dopasowania]) lub też istnieje konieczność zmiany funkcji oceny w trakcie rozwiązywania problemu (pojawiają się nowe agrumenty).

Problem

Dlaczego niektóre problemy są trudne do rozwiązania?

1. Niektóre problemy są skomplikowane (posiadają wiele zmiennych wpływających na wynik i przebieg rozwiązania), i nawet jeśli znamy metodę ich rozwiązywania, to konieczne jest ich uproszczenie (zmniejszenie liczby argumentów [zmiennych w problemie] bądź zmniejszenie ograniczeń [relaksacja]), co prowadzić może do tego, że uzyskane rozwiązania są bezużyteczne.
2. Rozmiar przestrzeni rozwiązań (liczba rozwiązań dopuszczalnych) jest tak rozległy, że rozwiązanie poprzez sprawdzenie wszystkich możliwości jest nieefektywne obliczeniowo (czasowo bądź pamięciowo).
3. Liczba rozwiązań dopuszczalnych jest tak mała (ze względu na warunki [ograniczenia] zadania), że trudno jest wygenerować rozwiązanie w ogóle, nie mówiąc o uzyskaniu rozwiązania optymalnego.
4. Sposób oceny rozwiązania jest niepoprawny, nieprecyzyjny (niewłaściwie określona funkcja celu [kosztu, nagrody, dopasowania]) lub też istnieje konieczność zmiany funkcji oceny w trakcie rozwiązywania problemu (pojawiają się nowe argumenty).
5. Posiadamy niekompletną wiedzę o problemie lub nie posiadamy odpowiedniej metody rozwiązania (takiej, która dawałaby akceptowalny wynik w rozsądnym czasie).

Problem

ad.1. Rozwiązujemy model problemu. Chyba że problem jest prosty i uwzględnienie wszystkich argumentów) pozwala na uzyskanie jego rozwiązania przy akceptowalnej złożoności obliczeniowej.

ad.2. Jedynie problemy klasy P możemy rozwiązywać w ten sposób, choć czy na pewno dla dowolnie dużej instancji (z punktu widzenia określonego rozsądnego czasu)?

ad.3. Znalezienie rozwiązania dopuszczalnego ze względu na warunki określone w zadaniu może być obliczeniowo trudne.

ad.4. Funkcja oceny może nie uwzględniać odpowiedniej liczby argumentów lub przypisywać im niewłaściwą wagę (określenie istotności) [istotność jest przeszacowana bądź niedoszacowana]. Wtedy błędnie wskazuje właściwy kierunek przeszukiwania przestrzeni rozwiązań.

ad.5. Im mniej wiemy o problemie bądź metodach jego rozwiązywania, tym mniejsze szanse na znalezienie rozwiązania w ogóle, nie mówiąc o rozwiązaniu optymalnym ☺

Reprezentacja problemu

Problem istnieje w rzeczywistości i charakteryzują go pewne argumenty (różnie istotne dla jakości rozwiązania w zależności od celu). Rzadko możemy rozwiązać problem w rozsądnym czasie biorąc pod uwagę wszystkie argumenty, jedynie modelujemy problem korzystając z części argumentów (zmiennych). Z drugiej strony pewne elementy świata rzeczywistego nie muszą być brane pod uwagę ze względu na ich nieistotność dla rozwiązania.

…

Korzystamy z pewnej reprezentacji problemu charakteryzowanej przez:

- stany, które reprezentują opis różnych stanów świata modelowanego,
- akcje, które reprezentują działania zmieniające stan problemu (tzw. operatory),
- koszt, który reprezentuje koszt związany z wykonaniem akcji.

Sformułowanie problemu

W najprostszy sposób problem możemy sformułować przez określenie:

- stanu początkowego – początkowe wartości argumentów charakteryzujących problem,
- stanu końcowego – określenie warunków osiągnięcia celu; określone wartości argumentów lub pewna formuła oceniająca jakość rozwiązania,
- rozwiązania – ciąg akcji prowadzący od stanu początkowego do stanu końcowego,
- kosztu rozwiązania* – określonego przez funkcję kosztu (celu, przystosowania); np. sumę wag krawędzi cyklu Hamiltona dla TSP.

* jeżeli problem jest problemem minimalizacyjnym szukamy rozwiązania o minimalnej wartości funkcji celu, jeśli maksymalizacyjny, to o największej wartości funkcji celu

Przykład problemu

Jak nie wysadzić w powietrze całego kwartału ulic, czyli zagadka z Die Hard with a Vengeance?

Water jug problem

Przykład problemu – water jug problem

Do dyspozycji mamy dwa naczynia:

nieprzezroczyste, o nieregularnym kształcie, bez wskaźnika pojemności,

o pojemnościach: 3 i 4 litry (wody).

…

Zadanie polega na odmierzeniu dokładnie 2 litrów wody i umieszczeniu ich w naczyniu 4-litrowym

…

Mamy określone ograniczenia:

naczynia można wypełnić lub uzupełnić do wartości maksymalnej (nalać do pełna) lub opróżnić – nie można dolać ani odlać trochę wody; można przelać część wody z naczynia do naczynia do wypełnienia drugiego.

Przykład problemu – WJP - argumenty

Naczynia w rzeczywistości posiadają kolor, są wykonane z jakiegoś materiału, mają określone pojemności, kształt, wagę, zapach i inne.

…

Które z ww., z punktu widzenia celu, są istotne i jak bardzo (istotność, ważność)?

Jak jest opisany stan problemu? W szczególności stan początkowy i końcowy.

Co jest rozwiązaniem problemu? Co otrzymamy rozwiązując problem? (oprócz braku eksplozji)

Jak oceniać wartość rozwiązania? Jak wygląda funkcja celu?

Czy można oceniać rozwiązania cząstkowe i czy mogą one pomóc w rozwiązaniu problemu? Jeśli tak, to w jaki sposób? Czy wartości rozwiązań cząstkowych mogą wpływać na wybór kolejnych akcji?

Czy otrzymane rozwiązanie może być przydatne w rozwiązywaniu podobnego (np. innego WJP z naczyniami 3 i 5 L) problemu? Jeśli tak, to jak, jeśli nie, to dlaczego?

Przykład problemu – WJP

Jak długo będziemy poszukiwać rozwiązania i jakiej będzie ono jakości, czyli siła metody i posiadane zasoby?

Przykład problemu – WJP

Jak długo będziemy poszukiwać rozwiązania i jakiej będzie ono jakości, czyli siła metody i posiadane zasoby?

aż skończy się dostępny czas – losowo lub brute-force dla TSP,

aż skończy się pamięć – algorytm Helda-Karpa (programowanie dynamiczne) dla TSP,

uznamy, że rozwiązanie jest odpowiednio dobre (satysfakcjonujące), np. gorsze od OPT o 10% [w przypadku problemów, dla których znamy OPT ☺] lub gorsze 15% od dolnego ograniczeni (lower-bound) [jeśli nie znamy OPT i potrafimy LB wyznaczyć],

uznamy, że algorytm niczego nie poprawił w ciągu ostatnich k iteracji bądź w ostatnich t sekundach – określamy w ten sposób warunek zatrzymania (jeden ze sposobów dla TS, SA, GA i ACO),

otrzymamy rozwiązanie optymalne – dla problemów klasy NP, jedynie dla małych instancji będziemy mieli na to szanse.

Rozwiązanie

Czego szukamy?

Rozwiązanie

Czego szukamy?

Ścieżki w przestrzeni stanów.

Ścieżki (drogi) w grafie, w którym węzłami są stany, a krawędzie ruchami (operatorami). Graf może być ważony.

Zestawu ruchów prowadzących od stanu początkowego do stanu końcowego.

…

Procedury ratującej ludność fragmentu Nowego Jorku ☺

Rozwiązanie

Czym są ruchy?

Rozwiązanie

Czym są ruchy?

Zależy od problemu.

W WJP to działania polegające na przelewaniu wody: nalej (do pełna) do n1 lub n2, opróżnij n1 lub n2, dolej (do pełna) do n1 lub n2, przelej z n1 do n2 do wypełnienia n2 lub odwrotnie.

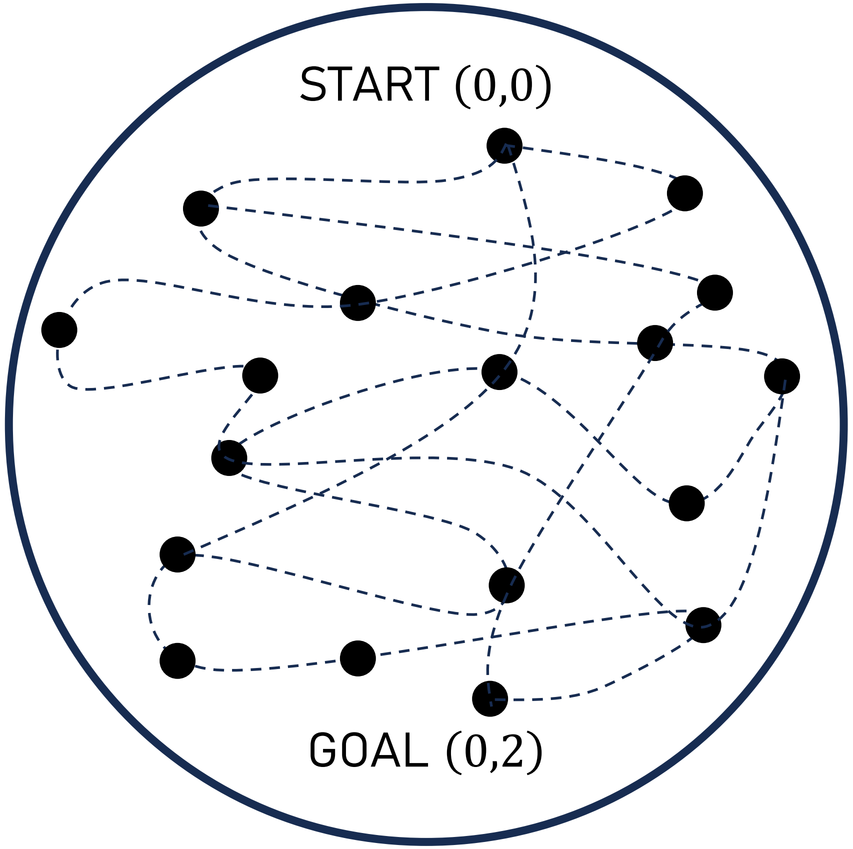
…

W TSP to sposoby wyboru kolejnych miast (węzłów) uzupełniające permutacje.

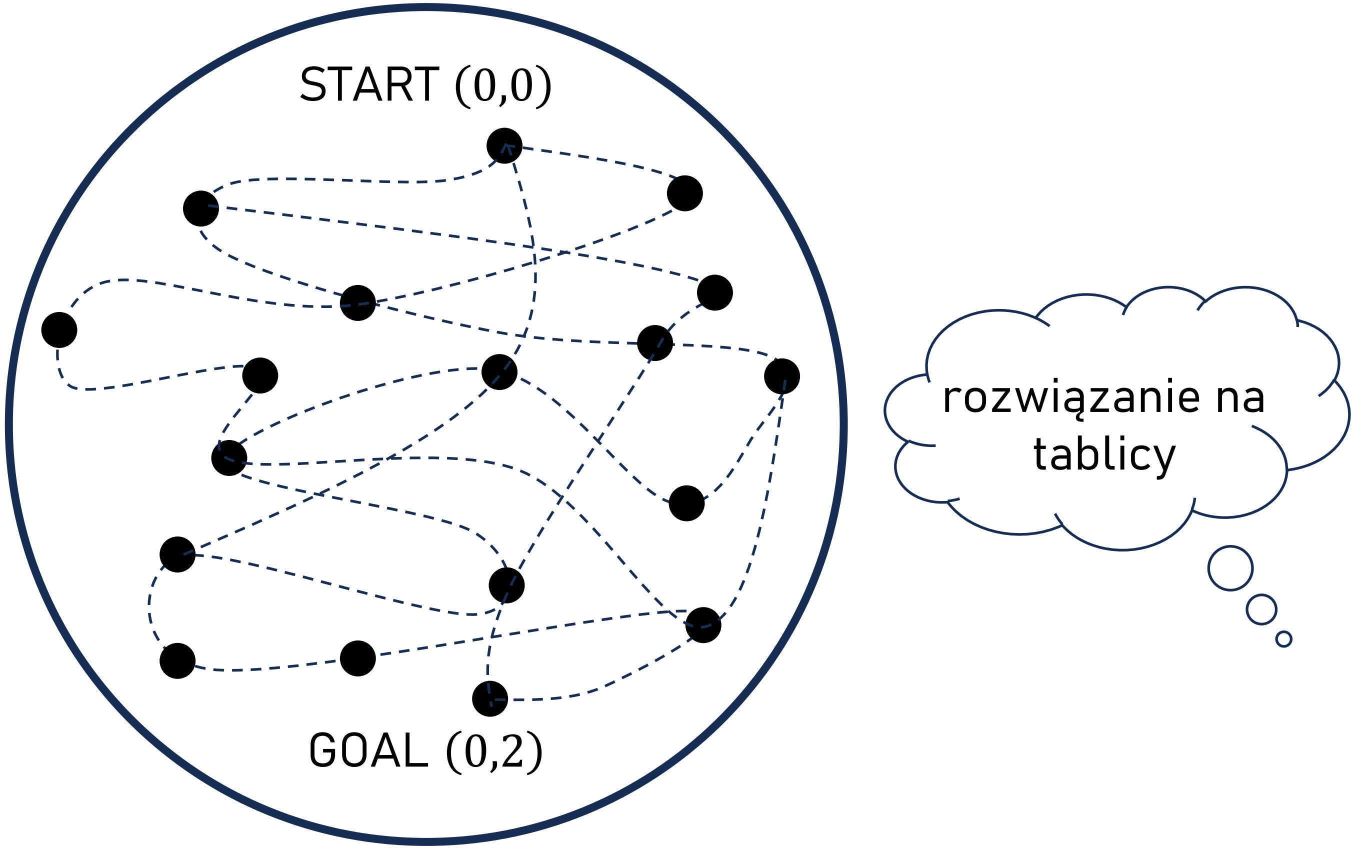
…

w 0/1 KSP to sposoby wyboru elementów wkładanych do plecaka, zgodnie z określonym kryterium (np. im cięższy obiekt (paczka) tym niżej musi być umieszczony w kontenerze.

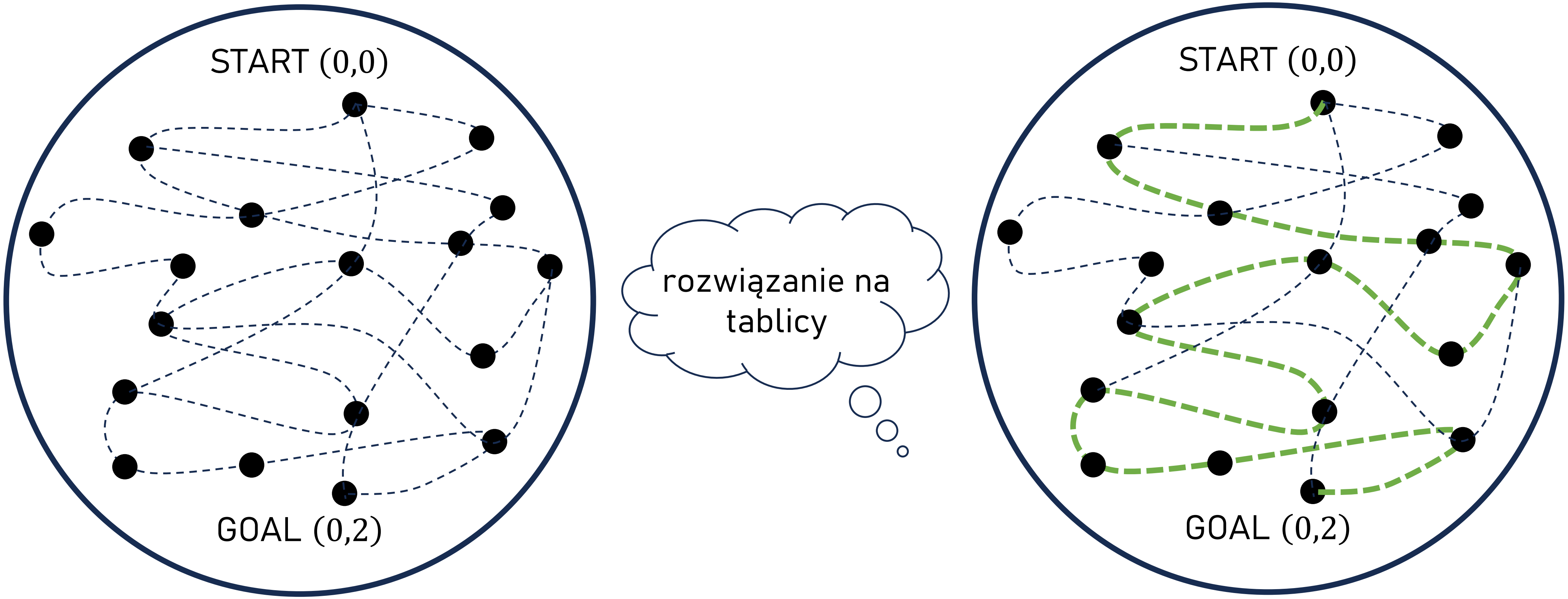
Rozwiązanie



Rozwiązanie



Rozwiązanie



Rozwiązanie

Jak wygląda rozwiązanie?

Jaka metoda (przeszukiwania przestrzeni) została zastosowana?

Jak wygląda algorytm? Czy w ogóle powstał?

Drobna uwaga dotycząca metod (strategii) przeszukiwania przestrzeni

Metoda (strategia) jest definiowana przez wybór kolejności ekspansji stanów.

Oceniane są wg następujących kryteriów:

- zupełność, czyli czy zawsze znajduje rozwiązanie, jeśli ono istnieje?
- optymalność, czyli czy znajduje rozwiązanie o minimalnym koszcie?
- złożoność czasowa, czyli wg liczby wygenerowanych węzłów,
- złożoność pamięciowa, czyli jaką maksymalną liczbę węzłów przechowuje w pamięci.

Drobna uwaga dotycząca metod (strategii) przeszukiwania przestrzeni

Złożoność obliczeniowa określane są przez:

b – maksymalne rozgałęzienie drzewa poszukiwań,

d – głębokość rozwiązania o najmniejszym koszcie,

m – maksymalna głębokość drzewa poszukiwań (może być ∞)

Drobna uwaga dotycząca metod (strategii) przeszukiwania przestrzeni

Strategie można podzielić na:

ślepe – korzystające jedynie z informacji zawartej w definicji problemu (nie korzystające np. z wag, czy sum krawędzi ścieżek częściowych do podejmowania decyzji o wyborze kierunku poszukiwania): np. wszerz, w głąb, ograniczone w głąb, z iteracyjnym pogłębianiem, dwukierunkowe, jednolitego kosztu.

…

heurystyczne – korzystające z dodatkowej, heurystycznej (przewidującej) funkcji oceny stanu (np. oceny kosztu dotarcia do stanu końcowego wg metryki euklidesowej (wyjście z lasu, z mapą): np. zachłanne, A*, lokalnie zachłanne

Drobna uwaga dotycząca metod (strategii) przeszukiwania przestrzeni

Przeszukiwanie wszerz (na przykład):

zupełność – tak, jeśli b jest skończone

optymalność – tak, jeśli koszt ruchu $= 1$ (generalnie nie)

złożoność czasowa - $\mathcal{O}\left(b^{d+1}\right)$

złożoność pamięciowa - $\mathcal{O}\left(b^{d+1}\right)$

Problemy, o których będziemy mówić

Problem komiwojażera (Travelling Salesmam Problem) – TSP	dużo i często
Problem plecakowy (Knapsack Problem) – dyskretny 0/1 KSP, ciągły KSP	mniej, choć często
Problem podziału (Partition Problem)	rzadko
Problem spełnialności formuł boolowskich (Boolean Satisfiability Problem) – SAT	j.w. ☺

TSP - definicja

Dane są zbiór n miast: $N = \{1, \dots, n\}$ oraz macierz odległości pomiędzy nimi $D = \{d_{ij}, i \in N, j \in N, i \neq j\}$, gdzie $d_{ij} \geq 0$ jest odległością z miasta i do miasta j. W ogólności $d_{ij} \neq d_{ji}$.

Polecenie: Znaleźć kolejność odwiedzania miast (permutację) $\sigma = \langle \sigma(1), \dots, \sigma(n) \rangle$, gdzie $\sigma(j)$ jest miastem odwiedzanym jako j-te, taką że

$$\sum_{i=1}^{n-1} d_{\sigma(i)\sigma(i+1)} + d_{\sigma(n)\sigma(1)} \to min$$

KSP - definicja

Dany jest zbiór n elementów $N = \{1, \dots, n\}$. Dla każdego elementu $j \in J$ określony jest jego rozmiar $a_j > 0$ oraz wartość $w_j > 0$. Dodatkowo dana jest pojemność plecaka $B > 0$.

Polecenie: Znaleźć podzbiór $X \subseteq N$ taki, że

$$\sum_{j \in J} a_j \leq B$$

oraz

$$\sum_{j \in J} w_j \to max$$

PP – definicja

Dany jest zbiór m elementów $N = \{1, \dots, m\}$ o wartościach $x_j > 0$, $j \in N$, taki, że $\sum_{j=1}^{m} x_j = 2B$.

Pytanie: Czy istnieje podzbiór $X \in N$ taki, że $\sum_{j \in X} q_j = B$?

SAT - definicja

Dana jest funkcja boolowska $f\colon\{0,1\}^n \to \{0,1\}$ ($f(x_1,\dots,x_n)$ jest funkcją boolowską zmiennych logicznych $x_1,\dots,x_n$)

Pytanie: Czy istnieje przyporządkowanie wartości 0 i 1 (logicznego fałszu i prawdy) do zmiennych $x_1,\dots,x_n$ takie, że $f(x_1,\dots,x_n) = 1$?

Pytania

Co wpływa i w jaki sposób na złożoność obliczeniową i jakość rozwiązania problemu?

Jak dobrać metodę do problemu?

Kiedy metoda (algorytm) można uznać za efektywny, czyli jakie są miary oceny efektywności algorytmów?

Jak bardzo model może być uproszczeniem problemu rzeczywistego?

Do przeczytania (minimum)

Z. Michalewicz, D. Fogel,

Jak to rozwiązać, czyli nowoczesna heurystyka

strony 35 - 64

Następnym razem

algorytmy dokładne; heurystyki i algorytmy heurystyczne