AUTOMATYKA • 2009 • Tom 13 • Zeszyt 2

Paweł Obszarski*, Marek Kubale*

Szeregowanie zadań wieloprocesorowych metodą kolorowania hiperkrawędzi

1. Wstęp

Od połowy ubiegłego stulecia teoria szeregowania zadań święci swój burzliwy rozwój, co spowodowane jest powiększającym się zapotrzebowaniem przemysłu, transportu i wielu innych branż na optymalne harmonogramy. Ostatnimi czasy obserwujemy dwie tendencje w działalności człowieka i maszyn przez człowieka tworzonych. Pierwszą jest specjalizacja. Wobec rosnącej wiedzy i skomplikowania świata, niemożliwym stało się, by jedna osoba mogła wiedzieć i robić wszystko. Podobnie jest z maszynami, okazuje się maszyny wyspecjalizowane są tańsze i lepiej niż maszyny uniwersalne wykonują swoje zadania. Druga tendencja to wieloprocesorowość, którą inaczej możemy nazwać pracą zespołową. Efekt synergii szczególnie widoczny jest w zespołach ludzkich, gdzie współpraca wielu osób pozwala osiągnąć wartość dodaną. Widzimy, jak te dwie tendencje wzajemnie się wzmacniają. Specjalizacja wymusza tworzenie zespołów maszyn lub ludzi o różnych kwalifikacjach i analogicznie – wieloprocesorowość pozwala na specjalizację.

Szeregowanie zadań, jako dziedzina opisująca działania człowieka, musi dostosowywać się do powyższych trendów, tak by możliwe było modelowanie nowych systemów. Dlatego w tej pracy rozważamy model szeregowania zadań wieloprocesorowych na procesorach dedykowanych z różnymi kryteriami optymalizacji szeregowania. Staramy się ustalić status złożoności obliczeniowej różnych typów instancji dla różnych kryteriów. Sprawdzamy, którędy przebiega granica między problemami wielomianowymi i NP-zupełnymi. I tak na przykład pokazujemy, że problem listowo-kosztowy jest wielomianowy dla hiperdrzew prostych, ale staje się trudny nawet dla 2-wierzchołkowych hiperdrzew, jeśli porzucimy ograniczenie liniowości (tj. ograniczenie, że dowolne dwie krawędzie mają co najwyżej jeden wspólny wierzchołek).

Jako model teoretyczny wiernie obrazujący ten problem wykorzystujemy problem kolorowania krawędzi hipergrafów.

* Katedra Algorytmów i Modelowania Systemów, Politechnika Gdańska

2. Definicja hipergrafów

Hipergrafem H nazywamy uporządkowaną parę $H = (V, E)$, gdzie V (lub $V(H)$) oznacza zbiór, a E (lub $E(H)$) reprezentuje multizbiór *krawędzi* (*hiperkrawędzi*). Elementy zbioru E są podzbiorami zbioru wierzchołków. Jeśli wiadomo, że krawędź zawiera k wierzchołków, to nazywać ją będziemy *k-krawędzią* lub *k-hiperkrawędzią*. W niniejszym opracowaniu przyjmujemy konwencję, że n (lub $n(H)$) oznacza liczbę wierzchołków hipergrafu lub grafu, a m (lub $m(H)$) – liczbę krawędzi. Zauważmy, że $E(H)$ zdefiniowane jest jako multizbiór krawędzi, zatem model dopuszcza występowanie wielokrotnych krawędzi. O dwóch krawędziach e_1 i e_2 będziemy mówić, że są *sąsiednie*, jeśli istnieje przynajmniej jeden wierzchołek v wspólny dla e_1 i e_2, czyli $\exists_{x\in V(H)} v\in e_1 \wedge v\in e_2$.

Dla hipergrafów możemy zdefiniować kolorowanie krawędzi, analogicznie jak dla grafów i multigrafów. *Poprawnym krawędziowym k-pokolorowaniem* (lub *pokolorowaniem*) hipergrafu H nazywamy funkcję c: $E(H) \to \{1, ..., k\}$ taką, że krawędzie mające wspólny wierzchołek otrzymują różne numery, zwane dalej *kolorami*. Z punktu widzenia kolorowania i szeregowania obecność 0-krawędzi, czyli krawędzi niezawierających wierzchołków, jest bez znaczenia, ponieważ krawędzie takie są odizolowane i nie interferują z pozostałymi krawędziami. Dla uproszczenia będziemy przyjmować, że hipergraf nie zawiera 0-krawędzi. Definicję kolorowania możemy rozszerzyć o listy dostępnych kolorów. Jeśli każdej hiperkrawędzi $e \in E$ przypisana jest skończona lista dostępnych kolorów $L(e)$ i poprawne pokolorowanie krawędziowe spełnia dodatkowy warunek $\forall_{e\in E} c(e) \in L(e)$, to będziemy mówili o *listowym pokolorowaniu krawędzi* hipergrafu. Skoro definicje kolorowania krawędzi są analogiczne jak dla grafów i multigrafów, to definicje indeksu chromatycznego oraz sumy chromatycznej również będą podobne. Minimalną liczbę kolorów potrzebnych, by poprawnie pokolorować krawędziowo hipergraf H nazywamy *indeksem chromatycznym* i oznaczamy $\chi'(H)$. *Optymalnym sumacyjnym pokolorowaniem krawędzi* hipergrafu H nazywamy funkcję c: $E(H) \to \{1, ..., k\}$ taką, że krawędzie mające wspólny wierzchołek otrzymują różne kolory oraz że *suma chromatyczna* $\Sigma'(H) = \sum_{e\in E(H)} c(e)$ jest minimalna. Najbardziej skomplikowane i najbardziej wymagające obliczeniowo modele to kosztowe i listowo-kosztowe kolorowanie hipergrafów. Rozważmy funkcję f_e: $IN \to IN$ (f_e: $L(e) \to IN$). Jeśli dla każdej krawędzi $e \in E$ z wykorzystaniem koloru $x \in IN$ ($x \in L(e)$) związany jest pewien koszt wyrażony funkcją $f_e(x)$, to kolorowanie będziemy nazywać *kosztowym* (*listowo-kosztowym*). W takiej sytuacji minimalizować będziemy sumę kosztów.

Istnieje wiele różnych klas hipergrafów. Większość z nich to odpowiedniki znanych klas grafów i multigrafów. Na wstępie przedstawimy klasę hipergrafów, która stanowi uogólnienie grafów prostych. *Hipergraf liniowy* to taki hipergraf, w którym każde dwie krawędzie mają co najwyżej jeden wspólny wierzchołek, czyli taki, że $\forall_{ei,ej\in E;ei\neq ej} |e_i \cap e_j| \leq 1$. *Hipergraf k-jednorodny* to hipergraf, w którym wszystkie krawędzie są k-elementowe. Zgodnie z powyższymi definicjami o multigrafach możemy powiedzieć, że są to hipergrafy 2-jednorodne. Grafy proste to 2-jednorodne hipergrafy liniowe. O hipergrafie H mówimy, że jest *indukowany* przez pewien graf prosty T, jeśli T ma ten sam zbiór wierzchołków co H

oraz każda krawędź $e \in E(H)$ (rozumiana jako pewien podzbiór zbioru $V(H)$) indukuje spójny podgraf w T. Graf T będziemy nazywać *grafem indukującym*. Za pomocą definicji hipergrafów indukowanych zdefiniujemy różne klasy hipergrafów oparte na pewnych strukturach grafowych. *Hiperlasem H* nazywamy taki hipergraf, że istnieje pewne drzewo T (graf prosty), które indukuje H. Spójny hiperlas nazywać będziemy *hiperdrzewem*. Analogicznie definiujemy hiperścieżki, hipercykle i hiperkaktusy. *Hiperścieżką* (*hipercyklem*, *hiperkaktusem*) H nazywamy taki hipergraf, że istnieje ścieżka (cykl, kaktus) T, która indukuje H. Istnieje jeszcze jedna klasa hipergrafów o drzewiastym charakterze, są to k-hiperdrzewa. *k-hiperdrzewem* nazywamy hipergraf, który po usunięciu co najwyżej k krawędzi staje się hiperdrzewem liniowym. Na koniec definiujemy prostą, acz bardzo przydatną do testowania pewnych własności, klasę hipergrafów. Hipergrafem *przecinającym się* nazywamy hipergraf H, w którym każde dwie krawędzie są sąsiednie, czyli $\forall_{ei,ej\in E(H)} e_i \cap e_j \neq \emptyset$.

3. Szeregowanie zadań

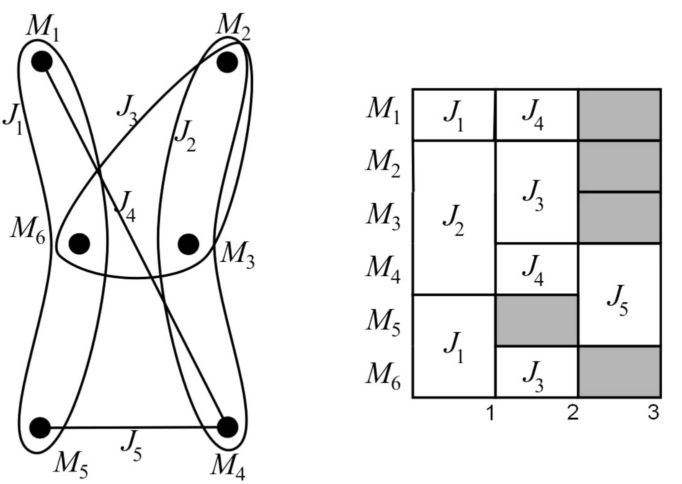
Rozważmy zbiór zadań $J = \{J_1, J_2, ..., J_m\}$, które mają być wykonywane na pewnym zbiorze procesorów $M = \{M_1, M_2, ..., M_n\}$ (zwanych także maszynami). Procesory różnią się między sobą i każdy jest wyspecjalizowany w unikalnej czynności. Aby wykonać zadanie J_i, konieczny jest pewien zbiór procesorów $fix_i \subseteq M$ dostępnych jednocześnie. Żaden procesor nie może pracować nad więcej niż jednym zadaniem w tym samym momencie. Czas jest podzielony na ponumerowane kolejnymi liczbami naturalnymi jednostki, które utożsamiać będziemy ze szczelinami czasowymi. Wszystkie zadania potrzebują jednakowej ilości czasu na wykonanie, dla uproszczenia możemy przyjąć, że wystarcza im jedna jednostka (szczelina) czasu, $p_i = 1$. Rozważać będziemy model z repetycją zadań, czyli taki, w którym zadania mogą się powtarzać. Zadania mogą mieć ograniczenia dostępności wyrażone skończonymi listami $L(J_i) \subset IN$ szczelin czasowych, w których dane zadanie może być wykonane. Kryteria optymalizacji mogą być różne, najbardziej klasyczne są *kryterium długości harmonogramu* $C_{\max}$ i *kryterium sumy czasów zakończenia zadań* ΣC_j. Dla zadania J_i jego wykonanie w pewnej szczelinie czasowej $x \in L(J_i)$ może wiązać się z kosztem określonym przez funkcję $f_{Ji}(x)$. W takiej sytuacji będziemy mówić o kryterium minimalizacji kosztu całkowitego.

Powyższy model będziemy nazywać szeregowaniem wieloprocesorowych zadań jednostkowych z ograniczeniami czasowymi na procesorach dedykowanych z kryterium kosztu całkowitego lub w skrócie szeregowaniem *listowo-kosztowym*. W notacji trójpolowej model listowo-kosztowy zapisujemy $P|p_j = 1, win, fix_j|\Sigma G_j(C_j)$. Symbol $\Sigma G_j(C_j)$ reprezentuje kryterium kosztu całkowitego. Model listowy z kryterium długości harmonogramu zapisujemy $P|p_j = 1, win, fix_j|C_{\max}$ zaś model bez list $P|p_j = 1, fix_j|C_{\max}$. Widać zatem, że symbol *win* w tym zapisie oznacza, że na zadania nałożone są ograniczenia dostępności (okna czasowe). Podobnie model z kryterium sumy czasów zakończenia z listami lub bez zapisujemy odpowiednio $P|p_j = 1, win, fix_j|\Sigma C_j$ i $P|p_j = 1, fix_j|\Sigma C_j$.

4. Kolorowanie a szeregowanie

Problemy szeregowania zadań dedykowanych i kolorowania krawędzi hipergrafów są równoważne. Maszyny utożsamiać będziemy z wierzchołkami, zadania z krawędziami, a kolory z ich terminami wykonania. Automatycznie, poprawne k-pokolorowanie krawędzi reprezentuje pewne uszeregowanie. Minimalna długość harmonogramu odpowiada indeksowi chromatycznemu $\chi'(H)$, a optymalne uszeregowanie względem kryterium $C_{\max}$ jest równoważne pokolorowaniu zużywającemu najmniejszą liczbę kolorów. Podobnie, uszeregowanie minimalizujące średni czas przepływu można utożsamiać z takim pokolorowaniem, które minimalizuje sumę kolorów $\Sigma'(H)$. Listy kolorów na krawędziach reprezentują listy dostępnych szczelin czasowych, analogicznie funkcje kosztu w szeregowaniu i kolorowaniu krawędzi są tożsame. Listowe kolorowanie hipergrafów utożsamiamy z szeregowaniem zadań z ograniczeniami czasowymi. Kosztowe i listowo-kosztowe kolorowanie hipergrafów reprezentuje odpowiednio szeregowanie zadań z kryterium minimalnego kosztu całkowitego i szeregowanie zadań z ograniczeniami czasowymi z kryterium kosztu całkowitego.

Hipergraf stworzony dla zobrazowania pewnego szeregowania zadań nazywać będziemy *hipergrafem szeregowania*. Hipergraf ilustrujący przykładową instancję problemu szeregowania zaprezentowaliśmy na rysunku 1. Od tej pory pojęcia szeregowanie zadań i kolorowanie krawędzi będą używane zamiennie.



Rys. 1. Przykład pewnej instancji problemu szeregowania zilustrowanej hipergrafem, oraz odpowiadającego jej wykresu Gantta

Różne klasy hipergrafów możemy interpretować jako pewne instancje problemu szeregowania. Hipergrafy planarne (hipergrafy, które można narysować na płaszczyźnie tak, by krzywe figury odpowiadające krawędziom nie przecinały się) możemy rozumieć jako procesory rozmieszczone na płytce, w taki sposób, że możliwy jest fizycznie nieprzerwany kontakt między procesorami potrzebnymi do wykonania zadania. Hipergrafy k-jednorodne reprezentują instancje szeregowania, w których wszystkie zadania są k-procesorowe. Graf indukujący pewien hipergraf H możemy interpretować jako pewną relację między procesorami taką, że tylko grupa procesorów, która tworzy spójną sieć relacji, może być wykorzystywana do wspólnego wykonania jednego zadania.

5. Złożoność problemów szeregowania

W niniejszym punkcie przedstawimy kilka podstawowych i kluczowych twierdzeń. Pozwolą one zaobserwować kilka tendencji.

5.1. Kryterium długości harmonogramu

Kryterium długości harmonogramu, w szczególności w wersji bez list, jest najprostsze z obliczeniowego punktu widzenia. Może być ono bez trudu sprowadzone do pozostałych kryteriów rozważanych w tym artykule. Efekt tego jest taki, że znamy relatywnie najwięcej klas hipergrafów, dla których problem ten jest wielomianowy.

Twierdzenie 1. *Istnieje wielomianowy algorytm o złożoności* $O(m)$ *dla kolorowania hiperdrzew względem kryterium indeksu chromatycznego.*

Dowód tego twierdzenia razem z algorytmem zaprezentowano w [8].

Dodanie list znacznie komplikuje problem, mamy bowiem

Twierdzenie 2. *Listowe kolorowanie krawędzi hiperścieżek jest NP-trudne, nawet jeśli listy krawędzi równych pod względem wierzchołków są takie same oraz krawędzie albo zawierają wszystkie wierzchołki, albo zawierają tylko jeden wierzchołek.*

Dowód tego twierdzenia zainspirowany [6] zaprezentowano w [7].

5.2. Kryterium sumy czasów zakończenia

Twierdzenie 3. *Problem ważonego sumacyjnego kolorowania krawędzi hipergrafu rozpiętego na dwóch wierzchołkach można rozwiązać w wielomianowym czasie* $O(m\log(m))$.

Dowód zaprezentowano w [7]. Twierdzenie uogólniające powyższe na dowolne hipergrafy k-wierzchołkowe można znaleźć w [5].

Twierdzenie 4. [3] *Problem sumacyjnego kolorowania krawędzi hiperdrzew jest NP-trudny już dla hiperdrzew o wymiarze* $\Psi(H) = 3$, *w których każda para hiperkrawędzi ma nie więcej niż dwa wspólne wierzchołki.*

5.3. Kryterium listowo-kosztowe

Kryterium kosztu całkowitego jest najtrudniejszym z rozpatrywanych przez nas. Tym niemniej udało się znaleźć kilka klas hipergrafów, dla których problem pozostaje wielomianowy.

Twierdzenie 5. *Problem listowo-kosztowego kolorowania krawędzi hipergrafów przecinających się można rozwiązać w czasie* $\mathrm{O}(m^3)$.

O trudności problemów z tym kryterium świadczy następne twierdzenie.

Twierdzenie 6. *Problem listowego kolorowania krawędzi hipergrafów jest NP-zupełny nawet dla hipergrafów dwuwierzchołkowych.*

Podobnie jak dowody następnego twierdzenia, dowód tego twierdzenia jest dostępny w [4] i w [7].

Twierdzenie 7. *Istnieje algorytm znajdujący optymalne listowo-kosztowe pokolorowanie hiperdrzewa liniowego w czasie* $\mathrm{O}(nN^2\log(nW))$, *gdzie* $N = N(T)$, $n = |V(T)|$, *a* W *to maksymalny koszt po wszystkich krawędziach i elementach ich list.*

W [4] pokazaliśmy również, że problem kolorowania k-hiperdrzew, które uogólniają hiperdrzewa, jest wielomianowy.

5.4. Zebrane wyniki

Wyniki badań nad problem szeregowania zadań i odpowiadającym mu problemem kolorowania krawędzi hipergrafów przedstawiamy w tabeli 1. Wiele pozycji z poniższej tabeli zostało wypełnionych zgodnie z zasadami:

- Jeśli problem ogólniejszy jest wielomianowy, to szczególny przypadek również.
- Jeśli przypadek szczególny jest NP-zupełny, to wersja ogólna zagadnienia także jest NP-zupełna.

Tabela 1
Status złożoności problemu szeregowania z różnymi kryteriami, różnych klas hipergrafów szeregowania

Nazwa	$C_{\max}$		$\Sigma w_j C_j$		$\Sigma G_j(C_j)$
	klasyczne	listowe	klasyczne	listowe	
hiperdrzewa liniowe	P	P	P	P	P
przecinające się	P	P	P	P	P
2-wierzchołkowe	P	?	P	?	NPC
hiperścieżki	P	NPC	NPC	NPC	NPC
hiperdrzewa	P	NPC	NPC	NPC	NPC
liniowe Δ=2	NPC	NPC	NPC	NPC	NPC

Zaobserwowaliśmy kilka ważnych zależności.

Po pierwsze szczególnie dla kryterium kosztu całkowitego wyraźna wydaje się być zależność między gęstością systemu a złożonością problemu. Gęstość rozumiemy tu jako ilość zadań w stosunku do ilości procesorów. Okazuje się, że problem jest NP-trudny nawet dla instancji dwuprocesorowych, podczas gdy jest łatwy obliczeniowo dla rzadkich hiperdrzew liniowych. Pewnym dysonansem odzywa się tu wynik, który mówi, że problem jest NP-trudny dla liniowych hipergrafów szeregowania stopnia 2, które są przecież systemami rzadkimi. Pamiętajmy jednak, że wierzchołki są dla tych systemów mnożone jakby w sztuczny sposób, tylko po to by osiągnąć $\Delta = 2$.

Po drugie potwierdzają się obserwacje innych autorów, że nałożenie ograniczeń dostępności zasobów znacząco pogarsza złożoność obliczeniową nawet dla problemów, które bez takich list są trywialne. Efekt ten mocno widoczny był przy badaniach nad kryterium długości harmonogramu.

Literatura

[1] Berge C., *Graphs and Hypergraphs*. Elsevier Science, 1985.

[2] Brandstadt A., Le V.B., Spinard J.P., *Graph Classes: A Survey.* SIAM Monographs on Discrete Mathematics and Applications, 1999.

[3] Giaro K., *Kolorowanie sumacyjne i listowo-kosztowe krawędzi hiperdrzew.* Raport badawczy na wydziale ETI Politechniki Gdańskiej, 2007.

[4] Giaro K., Kubale M., Obszarski P., *Graph coloring approach to scheduling of multiprocessor tasks on dedicated machines with availability constraints.* Discrete Applied Mathematics, (w druku).

[5] Krämer A., *Scheduling Multiprocessor Tasks on Dedicated Processors*. Dissertation Fachbereich Mathematik/Informatik, Universität Osnabrück, 1995.

[6] Marx D., *The complexity of tree multicolorings.* Lecture Notes in Computer Science 2420, 2002, 532–542.

[7] Obszarski P., *Szeregowanie zadań wieloprocesorowych na procesorach dedykowanych w modelu hipergrafowym*. Praca doktorska (w recenzji).

[8] Obszarski P., Dąbrowski J., *Hipergrafowy model szeregowania w rozrzedzonych systemach zadań wieloprocesorowych.* Zeszyty Naukowe Wydziału ETI Politechniki Gdańskiej, 10, 2006, 499–506.