

Wrocław, 26 stycznia 2022 r.

dr hab. Marcin Bieńkowski
Instytut Informatyki
Wydział Matematyki i Informatyki
Uniwersytet Wrocławski

Recenzja rozprawy habilitacyjnej
„Szeregowanie zadań w sieciach zbierających dane”
dr Joanny Berlińskiej

Joanna Berlińska ukończyła studia na Uniwersytecie im. Adama Mickiewicza (UAM) w Poznaniu i uzyskała tam również tytuł doktora nauk matematycznych w zakresie informatyki w 2011 roku na podstawie rozprawy „Scheduling divisible loads in heterogeneous distributed systems”. Od tego czasu pozostaje zatrudniona jako adiunktka na Wydziale Matematyki i Informatyki UAM.

Przedstawione przez nią osiągnięcie naukowe dotyczy algorytmów szeregowania przesyłu i przetwarzania danych w sieciach złożonych z grupy czujników zbierających dane i stacji bazowej. Rozważane są różne warianty takiego modelu z różnymi celami optymalizacyjnymi. Kandydatka konstruuje algorytmy, heurystyki, dowodzi ich poprawności i złożoności, przedstawia dowody trudności obliczeniowej, a także uzupełnia część teoretyczną symulacjami. W ramach swojego dorobku habilitantka wyróżniła poniższy cykl publikacji jako swoje osiągnięcie naukowe.

- (H1) Joanna Berlińska: Communication scheduling in data gathering networks with limited memory, *Applied Mathematics and Computation* (2014), 530-537.
- (H2) Joanna Berlińska: Scheduling for data gathering networks with data compression, *European Journal of Operational Research* (2015), 744-749.
- (H3) Joanna Berlińska: Scheduling data gathering with maximum lateness objective, *PPAM* 2017, 135-144.
- (H4) Joanna Berlińska: Scheduling in a data gathering network to minimize maximum lateness, *Operations Research Proceedings* 2018, 453-458.
- (H5) Joanna Berlińska: Makespan minimization in data gathering networks with dataset release times, *PPAM* 2019, 230-241.
- (H6) Joanna Berlińska: Heuristics for scheduling data gathering with limited base station memory, *Annals of Operations Research* (2020), 149-159.
- (H7) Joanna Berlińska: Scheduling in data gathering networks with background communications, *Journal of Scheduling* (2020) 681-691.

(H8) Joanna Berlińska: A comparison of priority rules for minimizing the maximum lateness in tree data gathering networks, Engineering Optimization (2021)

1. Omówienie i ocena wskazanego osiągnięcia habilitacyjnego

Rozprawa dotyczy algorytmiki dla problemów optymalizacyjnych dotyczących szeregowania. Typowo w takich problemach zadanie polega na rozdzieleniu zadań pomiędzy dostępne procesory, tak aby zoptymalizować pewną funkcję celu, np. czas zakończenia wykonywania wszystkich zadań. Jest to bardzo szeroka tematyka będąca w centrum zainteresowania informatyki, matematyki dyskretnej i badań operacyjnych. Większość spotykanych zagadnień jest obliczeniowo trudna i zazwyczaj rozważane są wtedy algorytmy przybliżone, badane zarówno w ujęciu teoretycznym (np. algorytmy aproksymacyjne), jak i w ujęciu eksperymentalnym, gdzie wydajność (pod względem funkcji celu) i czas działania algorytmów badana jest na instancjach pochodzących z praktycznych zastosowań.

W swojej rozprawie habilitacyjnej Joanna Berlińska koncentruje się na zagadnieniach szeregowania motywowanych zbieraniem pomiarów w systemach rozproszonych, takich jak sieci czujników wykonujących pomiary i pewne obliczenia. Zakłada się tutaj, że rozmiar zbieranych danych jest na tyle duży, że warto jest zoptymalizować czas ich przekazywania do centralnego punktu.

We wszystkich przedstawionych pracach (poza częściowym wyjątkiem pracy [H8]) rozważana jest topologia gwiazdy, gdzie m czujników jest bezpośrednio połączonych z pojedynczą stacją bazową. W trakcie działania, każdy z czujników pozyskuje pewną porcję danych o rozmiarze α_j , która gotowa jest do przesłania do stacji bazowej w czasie r_j . Następnie taka porcja może zostać przekazana do stacji bazowej w czasie, który jest afiniczną funkcją α_j , a potem przetworzona przez stację bazową w czasie, który jest liniową funkcją α_j . Zakładamy, że stacja bazowa może odbierać komunikaty od jednego czujnika naraz i przetwarzać jedną porcję danych naraz (ale pobieranie i przetwarzanie mogą odbywać się jednocześnie).

Minimalizacja czasu zakończenia

Z powyższego opisu wynika, że powyższe zagadnienie jest szczególnym przypadkiem tzw. problemu flow-shop scheduling, gdzie każde zadanie musi przejść przez dwa etapy przetwarzania (w tym przypadku: przesyłanie do stacji bazowej i przetwarzanie w niej). Szczególnym — bo czasy przetwarzania w pierwszym i drugim kroku są w rozważanym modelu zależne od siebie, podczas gdy w problemie flow-shop scheduling taka zależność nie istnieje. Wiadomo, że jeśli celem optymalizacji jest minimalizacja czasu zakończenia ostatniego zadania, to problem flow-shop scheduling $F_2|r_j|C_{\max}$ jest NP-trudny. Jeśli natomiast wszystkie porcje danych dostępne są od razu ($r_j = 0$ dla każdego j), to problem może zostać rozwiązany w czasie wielomianowym za pomocą tzw. reguły Johnsona.

Habilitantka pokazała, że problem ten pozostaje NP-trudny w rozważanym przez nią wariacie (gdy r_j są różne, a czasy przesyłania i przetwarzania są zależne od siebie). W pracy [H5] stworzyła też efektywne czasowo algorytmy 2-aproksymacyjne oparte o wykorzystanie reguły Johnsona. Dowód NP-trudności opiera się o rutynową redukcję z problemu 3-partition, analiza algorytmu też nie pokonuje istotnych trudności i jest dość elementarną analizą przypadków.



Wariant rozważanego problemu, w którym pamięć stacji bazowej mieści ograniczoną ilość danych została przedstawiona w pracy [H6]. Również w tym przypadku problem okazuje się NP-trudny (redukcja z problemu bin-packing), a autorka przedstawia prostą zachłanną heurystykę rozwiązującą sub-optimalnie ten problem.

W pracy [H7] autorka rozważa przypadek, w którym co prawda przetwarzanie danych w stacji bazowej nie jest konieczne, ale za to czas przesyłania danych może zmieniać się na skutek obciążenia łącz innymi transmisjami pozostającymi poza kontrolą algorytmu. Wersja z wywłaszczeniem jest rozwiązywalna w naturalny sposób za pomocą programowania liniowego, zaś wersja bez wywłaszczenia jest NP-trudna (redukcja z 3-partition). Autorka koncentruje się na efektywnych heurystykach które rozwiązują zadanie w przybliżony sposób; niestety osiągnięte przez nie współczynniki aproksymacji są takie jak osiągnięte przez trywialne podejścia.

Minimalizacja opóźnienia

Inną rozważaną w rozprawie klasą zagadnień jest przypadek kiedy na przetworzenie każdej porcji danych mamy określony deadline i chcemy zminimalizować jego przekroczenie (maksymalizowane po wszystkich porcjach danych). Autorka zajmuje się tym w cyklu prac [H3], [H4], [H8]. Wkład w dziedzinę jest tutaj podobny, jak w przypadku minimalizacji czasu zakończenia: dowiedziona jest NP-trudność i tworzone są szybkie i proste heurystyki, które analizowane są eksperymentalnie.

Są tutaj dwie pozytywne różnice, o których warto wspomnieć. Przede wszystkim praca [H3] pokazuje interesujący kombinatoryczny fakt, że w wariancie z wywłaszczeniem możliwe są instancje, dla których do osiągnięcia optymalnego wyniku konieczne jest wywłaszczenie w momencie, który jest różny od wszystkich momentów generowania porcji danych. W szczególności oznacza to, że algorytmy prezentowane w pracy [34] (Information Processing Letters 2002) są niepoprawne. Drugą różnicą jest praca [H8], która wychodzi poza schemat topologii gwiazdy i rozważa wariant dwupoziomowego drzewa.

Inne rozpatrywane warianty

W pracy [H1] autorka rozważa zagadnienie maksymalizacji czasu życia sieci, gdzie każdy z sensorów ma ograniczoną pamięć i zbiera dane w ciągły sposób. Celowe jest zatem zapewnienie, że dane z sensora zdążą być wysłane do stacji bazowej zanim pamięć sensora się przepełni. Habilitanka pokazała w dość prosty sposób, jak efektywnie wyliczać optymalne uszeregowanie wysyłania danych.

W pracy [H2] rozważany jest wariant, w którym dane przed wysyłaniem mogą być kompresowane: z jednej strony zmniejsza to czas ich przesyłania, z drugiej zaś generuje dodatkowy koszt (np. energetyczny). Celem jest dostarczenie wszystkich porcji danych do określonego czasu do stacji bazowej i jednoczesna minimalizacja kosztu kompresji. Autorka pokazała, że problem ten jest NP-trudny nawet w przypadku jednorodnym, w którym porcje danych sensorów są jednakowe i koszty przesyłu są takie same (różne są tylko czasy r_j zebrania danych przez czujniki). Tak jak w przypadku innych jej prac, algorytmiczna część koncentruje się na działającej szybko heurystyce, której jakość sprawdzana jest eksperymentalnie.

Ocena osiągnięcia habilitacyjnego

Habilitantka gruntownie przeanalizowała wiele podproblemów szeregowania zadań związanych z problemem zbierania danych. Rozważane warianty są osadzone w rzeczywistości i oparte na zastosowaniach: mają sensowną i przekonującą motywację w postaci zbierania danych z sieci czujników.

Z teoretycznego punktu badania kandydatki wpisują się w tradycyjny nurt analizy algorytmów: pokazywana jest trudność obliczeniowa problemu i następnie prezentowane są algorytmy aproksymacyjne dla tego problemu (bądź szybkie heurystyki bez gwarancji na jakość rozwiązania). Negatywnym aspektem jest to, że praktycznie wszystkie wyniki są dość rutynowe (choć nietrywialne). Dowody NP-trudności opierają się głównie na redukcji z problemu 3-partition, zaś analiza algorytmów aproksymacyjnych jest prostą analizą kilku przypadków. Tworzone algorytmy wykorzystują różne techniki (np. branch-and-bound, programowanie liniowe, programowanie dynamiczne, local search), ale techniki te wykorzystywane są instrumentalnie na etapie tworzenia algorytmów; późniejsza analiza ogranicza się zazwyczaj do łatwego oszacowania czasu wykonania. Nieskomplikowanie techniczne wyników jest raczej problematyczne biorąc uwagę, że rozważone problemy są postawione przez autorke. Na plus należy natomiast ocenić to, że niektóre ze zdefiniowanych przez nią wariantów spotkały się z zainteresowaniem i naukową kontynuacją przez innych badaczy, np. praca [36] (Algorithmica 2018).

Wszystkie prace zawierają symulacje prezentowanych i znanych wcześniej algorytmów. Jest to doskonale uzupełnienie pokazujące, że wyniki teoretyczne w praktyce są czasem zbyt pesymistyczne, czasem też będące jedynym źródłem informacji o ich jakości. Z drugiej strony trudno nie zauważyć, że wszystkie symulacje przeprowadzane są na syntetycznych danych pochodzących z losowych rozkładów. Rozprawa habilitacyjna istotnie by zyskała, gdyby przeprowadzono implementację algorytmów w rzeczywistym środowisku, lub chociaż zbadano ich efektywność na danych pochodzących z rzeczywistości.

Część uzyskanych w osiągnięciu habilitacyjnym wyników została opublikowana w dobrych miejscach. Na liście brakuje czołowych konferencji i czasopism, ale znajduje się na niej kilka dobrych bądź bardzo dobrych pozycji. W szczególności, praca [H2] została opublikowana w European Journal of Operational Research, zaś praca [H7] w Journal of Scheduling. Niezłym periodykiem jest też Annals of Operations Research, w którym ukazała się praca [H6]. Publikacje pokonferencyjne nie mają już tej jakości: na liście znajdują się wyłącznie konferencje odbywające się w jednym kraju takie jak PPAM (polska konferencja o międzynarodowych aspiracjach) oraz Operations Research Proceedings (niemiecki workshop).

Wszystkie osiem prac wchodzących w skład rozprawy prac zostało napisane samodzielnie. Z jednej strony zdecydowanie pozytywnie wynika z tego samodzielność kandydatki, z drugiej zaś otwarcie na współpracę z innymi naukowcami pomogłoby poszerzyć naukowe horyzonty i ułatwiłoby uzyskiwanie nowych i ciekawych wyników. Warto dodać, że w ramach pozostałego dorobku jest już parę prac z innymi naukowcami, również spoza Polski, choć grono współpracowników pozostaje kilkusobowe.

2. Ocena pozostałej aktywności naukowej, dydaktycznej i organizacyjnej

Pozostały dorobek kandydatki (poza osiągnięciem habilitacyjnym) dotyczy również szeregowania zadań, choć w innych ujęciach (tzw. divisible load, obliczenia MapReduce). W autoreferacie znajdujemy 17 takich pozycji, jednak część z nich jest na workshopach nienotowanych w DBLP; baza DBLP wymienia takich publikacji 10. Miejsca ich publikacji mają podobny prestiż do tych przedstawionych w osiągnięciu habilitacyjnym. Warto jednak dodać, że w ciągu ostatnich dwóch lat grono współpracowników habilitantki znacznie się poszerzyło.

Sumarycznie (razem z osiągnięciem habilitacyjnym) prace cytowane są ok. 250 razy wg. Google Scholar, 125 razy wg. AMiner, co jest dobrym wynikiem. Najlepiej cytowanymi pracami z dorobku habilitacyjnego są najwcześniejsze prace: [H2] dotycząca kompresji (23 cytowania, 11 bez auto-cytowań) i [H1] wprowadzająca model (18 cytowań, 7 bez auto-cytowań). Rozpoznawalność autorki w środowisku jest też odzwierciedlona w fakcie, że była członkiem komitetu programowego Euro-Par (dość istotna konferencja dotycząca systemów równoległych i rozproszonych).

Kandydatka brała udział w kilku projektach badawczych, jej badania zostały też docenione przez Narodowe Centrum Nauki: była kierownikiem grantu PRELUDIUM i (obecnie) grantu SONATA. Dość wysoko oceniam aktywność konferencyjną habilitantki: w latach 2016-2019 wygłosiła 11 referatów na różnych konferencjach i workshopach. Pewien niepokój budzi brak wyjazdów naukowych w ramach współpracy z innymi naukowcami, choć można to w pewnej mierze usprawiedliwiać pandemią COVID-19 (w szczególności uniemożliwiła ona zaplanowany na 2020 rok wyjazd do Uniwersytetu w Sydney).

Aktywność dydaktyczna i organizacyjna kandydatki jest odpowiednia do obecnego etapu kariery naukowej: prowadzi wykłady (z algorytmów i struktur danych), ćwiczenia i laboratoria; opiekuje się kilkoma magistrantami. Jest również członkiem Komitetu Okręgowego Olimpiady Matematycznej i prowadzi zajęcia dla międzyszkolnego koła matematycznego.

3. Konkluzja

Pomimo przedstawionych wyżej krytycznych uwag uważam, że przedstawiona rozprawa habilitacyjna i dorobek naukowy Joanny Berlińskiej spełniają ustawowe i zwyczajowe wymagania stawiane rozprawom habilitacyjnym. Cykl przedstawionych prac stanowi konsekwentną realizację spójnego przedsięwzięcia, jakim jest badanie problemów szeregowania zadań związanych ze zbieraniem danych w sieciach czujników i stanowi niebanalny wkład w dziedzinę szeregowania. **Popieram wniosek dr Joanny Berlińskiej i wnioskuję o jej dopuszczenie do dalszych etapów przewodu habilitacyjnego.**

Mariusz Reański