

## **Recenzja rozprawy doktorskiej mgr. Bartosza Rybickiego „Approximation algorithms for NP-hard location problems”**

### **Ocena**

Po przeanalizowaniu przedłożonej rozprawy doktorskiej niniejszym stwierdzam, że **spełnia ona wymagania** „Ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki” i wnioskuję o dopuszczenie mgr. Bartosza Rybickiego do dalszych etapów przewodu doktorskiego. Wnioskuję również o wyróżnienie rozprawy.

### **Uzasadnienie**

#### **Wstęp**

Rozprawa dotyczy algorytmów dla kilku wariantów problemu rozmieszczenia fabryk (ang. facility location problem). W problemach tych mamy dane zbiory klientów  $C$  i możliwych lokalizacji fabryk  $F$ , umieszczone w pewnej przestrzeni metrycznej. Celem jest wybór pewnego podzbioru  $F' \subseteq F$  (oznaczającego miejsca, w których otwieramy fabryki) oraz przypisanie każdego klienta do pewnej fabryki z  $F'$ . W zależności od wariantu problemu, należy spełnić rozmaite ograniczenia (ograniczona liczba fabryk, ograniczona liczba klientów przypisanych do jednej fabryki, itp.). Poszukiwane są rozwiązania optymalizujące pewną funkcję celu. W opisywanej pracy jest to niemal zawsze suma odległości klientów do przypisanych im fabryk.

Choć pojęcia „fabryka” i „klient” należy traktować umownie, opisana grupa problemów jest bardzo istotna z praktycznego punktu widzenia, gdyż modeluje wiele rzeczywistych problemów optymalizacyjnych (lokalizacja sieci restauracji, koncertów, przypisanie serwerów do komputerów-klientów). Dlatego już od ponad 30 lat problemy lokalizacji fabryk są w centrum zainteresowania badaczy. Z wyjątkiem kilku bardzo szczególnych przypadków problemy te są NP-trudne, co najprawdopodobniej wyklucza dokładne rozwiązania wielomianowe. Dlatego od lat 90-tych trwają badania nad algorytmami aproksymacyjnymi. Problemy lokalizacji fabryk zajmują szczególne miejsce w teorii algorytmów aproksymacyjnych. Z jednej strony, uzyskanie jakiegokolwiek stałego współczynnika aproksymacji jest wysoce nietrywialne: zachłanne algorytmy nie działają, udało się to dopiero w drugiej połowie lat 90-tych dzięki użyciu zaawansowanych metod opartych na programowaniu liniowym. Z drugiej strony, w wielu najbardziej naturalnych wariantach tego problemu wyścig o uzyskanie najlepszego współczynnika aproksymacji trwa do dziś, a biorą w nim udział czołowi badacze z najważniejszych światowych ośrodków informatyki teoretycznej. Można tu wymienić nazwiska takie jak Charikar (Princeton), Vazirani

(Georgia Tech), Svensson (EPFL). Pan Bartosz Rybicki w swoim doktoracie podejmuje wyzwanie konkurowania z najlepszymi, zajmując się trzema wybranymi wariantami. We wszystkich przypadkach proponowane rozwiązania to złożone algorytmy zaokrąglania rozwiązania ułamkowego programu liniowego, co nie dziwi, gdyż jest to jak dotąd wiodące w literaturze podejście do aproksymowania tych problemów.

## Wyniki rozprawy

Pierwszy rozważany wariant to Knapsack Median (KM), gdzie każda fabryka ma swój koszt otwarcia i nie należy przekroczyć pewnego budżetu  $B$ . Wcześniej problem był badany kolejno przez Kumara [SODA'12], Charikara i Li [ICALP 2012] oraz Swamy [APPROX 2014], którzy uzyskiwali coraz lepsze współczynniki aproksymacji. W rozdziale trzecim rozprawy opisano algorytm 17.46-aproksymacyjny, który poprawia współczynnik 32 z pracy Swamy. Podobnie jak we wcześniejszych pracach, stosuje się technikę zaokrąglania PL. Jest to jednak utrudnione, ze względu na nieograniczoną lukę całkowitości naturalnego programu liniowego. Kłopot ten rozwiązał Kumar, podając szacowanie na koszt połączenia skupionej blisko siebie grupy klientów w odniesieniu do kosztu rozwiązania całkowitoliczbowego (a nie ułamkowego). W rozdziale trzecim znajdujemy sprytne wzmocnienie szacowania Kumara dla tzw. rzadkich instancji, co w połączeniu z zaproponowaną niedawno techniką rozrzedzania Li i Svenssona [STOC'13] oraz twórczo przetworzonymi lub usprawnionymi pomysłami z wcześniejszych prac dało istotnie lepszy współczynnik aproksymacji.

Kolejne dwa rozdziały poświęcone są problemowi Capacitated  $k$ -Median, gdzie każda fabryka może obsłużyć jedynie  $u$  klientów, oraz mamy ograniczenie  $k$  na liczbę fabryk. Nie jest znany algorytm aproksymacyjny o stałym współczynniku dla tego problemu. Jednym z powodów tego stanu rzeczy jest nieograniczona luka całkowitości PL, i to nawet gdy ograniczenia  $u$  lub  $k$  mogą być przekroczone 2-krotnie. Głównym wynikiem rozdziału 4 jest algorytm o stałym współczynniku, który dla dowolnego  $\epsilon > 0$  przekracza ograniczenie na liczbę obsłużonych klientów co najwyżej  $(2 + \epsilon)$ -krotnie (nie przekraczając ograniczenia na liczbę fabryk), co wyczerpuje możliwości standardowego programu liniowego ze względu na wspomnianą lukę całkowitości. Współczynnik aproksymacji algorytmu zależy od  $\epsilon$  i wynosi  $O(\epsilon^{-2})$ . Punktem wyjścia były prace Charikara et al. [STOC'99] i Levi et al. [IPCO'04], w których klienci dzieleni są na klastry, w których możliwe jest efektywne znalezienie niemal całkowitoliczbowego rozwiązania. Niestety klastry nie mogą być rozwiązywane niezależnie i czasem klienci muszą szukać swoich fabryk w sąsiednich klastrach. Stąd główna trudność polegała na połączeniu wspomnianych lokalnych rozwiązań w rozwiązanie globalne, przez zbudowanie drzewiastej struktury opisującej klastry i odpowiednie przekierowywanie klientów w ramach tej struktury. Uzyskane rozwiązanie, choć skomplikowane, jest imponujące; korzysta z wielu ciekawych narzędzi (zaokrąglanie zależne, algorytmy przepływowe). W mojej ocenie jest to najważniejsza część rozprawy. Rozdział 4 zawiera też nieco słabszy wynik  $(3 + \epsilon)$  zamiast  $(2 + \epsilon)$  dla ogólniejszego wariantu, w którym każda fabryka może mieć własny limit na liczbę obsługiwanych klientów.

W 2014 r. Shi Li [SODA'15] nieoczekiwanie znalazł sposób na obejście luki całkowitości zwykłego programu liniowego dla problemu Capacitated  $k$ -Median. Zastosował on metodę *round-or-separate* (Carr et al. [SODA'00]), mianowicie użył nowego programu liniowego o wy-

kładniczej liczbie warunków. Pomimo dużych rozmiarów programu liniowego jego algorytm wykonuje wielomianowo wiele kroków: zaczyna od zwykłego programu liniowego próbując zaokrąglić rozwiązanie ułamkowe, a w przypadku niepowodzenia znajduje niespełniony warunek nowego programu i nowe rozwiązanie ułamkowe uwzględniające ten warunek. Li uzyskał współczynnik aproksymacji rzędu  $\exp(O(\epsilon^{-2}))$  przy co najwyżej  $(1 + \epsilon)k$  otwartych fabrykach. Rok później [SODA'16] Li poprawił współczynnik do  $O(\epsilon^{-2})$ , wprowadzając jeszcze skuteczniejszy program liniowy (tzw. konfiguracyjny PL). W rozdziale piątym znajdujemy przeniesienie techniki Li do sytuacji rozważanej w rozdziale 4, tzn. gdy dopuszczalne jest przekroczenie ograniczenia na liczbę klientów obsługiwanych przez fabrykę, a nie liczbę fabryk. Zachowując współczynnik aproksymacji  $O(\epsilon^{-2})$ , algorytm przypisuje każdej fabryce  $O(1 + \epsilon)u$  klientów. Choć widać tam wiele z pomysłów Li (konfiguracyjny PL, tzw. drzewa czarno-szaro-białe), ze względu na odmienne ograniczenia algorytm zaokrąglający musi być zupełnie inny, w szczególności udało się przenieść sporo pomysłów z rozdziału 4. Sądzę, że rozdział 5 jest również bardzo cenny, gdyż technika *round-or-separate* jest w dalszym ciągu mało znana i w niewielkim stopniu zbadana.

W ostatnim rozdziale autor skupia się na problemie  $k$ -poziomowej lokalizacji fabryk, w którym mamy  $k$  rozłącznych zbiorów (*poziomów*) fabryk, a dla każdego klienta należy znaleźć ścieżkę przez kolejne poziomy. Ponownie znajdujemy tu szereg ciekawych technik algorytmicznych, w tym sformułowanie nowego programu liniowego, klastrowanie klientów, randomizowane zaokrąglanie.

## Poziom redakcyjny rozprawy

Poziom redakcyjny jest akceptowalny, jest on jednak słabą stroną recenzowanej rozprawy. Najpoważniejsze zastrzeżenie dotyczy połączenia poszczególnych wyników (zawartych w kolejnych rozdziałach i pochodzących z osobnych artykułów mgr. Bartosza Rybickiego i jego współautorów) w całość. Niestety wydaje się, że treść tych rozdziałów niezbyt odbiega od oryginalnych artykułów. W konsekwencji zdarza się, że te same pojęcia są definiowane kilka razy, w dodatku z użyciem innych oznaczeń (przykład: ułamkowy koszt połączenia klienta). Rozdział 5 korzysta z niektórych pojęć i lematów sformułowanych w rozdziale 4, jednak zamiast odnośników do rozdziału 4 mamy odnośniki do oryginalnej pracy, na podstawie której rozdział 4 powstał (lub prac innych autorów). Doktorat powinien być okazją do zunifikowania systemów pojęciowych z poszczególnych prac, spojrzenia na te prace na nowo, co mogłoby zaowocować czystsza prezentacją. Szkoda, że autor nie skorzystał z tej okazji. Nowością w stosunku do materiału pochodzącego z oryginalnych prac jest kilkustronicowy wstęp na temat problemów lokalizacji fabryk oraz osobny rozdział poświęcony technice *round-or-separate* (jego pozycja w strukturze pracy jest nieco dziwna; bardziej naturalne byłoby gdyby był to fragment rozdziału 5 — jedyne, w którym ta technika jest używana).

W rozprawie znajdujemy szereg skomplikowanych, wielostronicowych konstrukcji, dość nieprzyjaznych dla czytelnika. Dużą pomocą dla czytelnika byłoby naszkicowanie, na początku każdego takiego opisu zarysu następującego dalej rozumowania, lub chociaż wytyczenie celu, który taka konstrukcja osiąga.

W tekście można wskazać szereg literówek oraz błędów językowych. Dużym utrudnieniem dla recenzenta była przypadkowa kolejność pozycji w bibliografii.

## **Podsumowanie**

Rozprawa doktorska mgr. Bartosza Rybickiego jest wynikiem zakończonym sukcesem bardzo ambitnego i tematycznie spójnego programu badawczego. Już samo dogłębne zrozumienie technik stosowanych we wcześniejszych pracach dotyczących rozważanych problemów jest dużym osiągnięciem, zaś ich twórcze łączenie i modyfikowanie świadczy o bardzo wysokich kompetencjach doktoranta. Wreszcie, duży szacunek budzi atakowanie naturalnych, studiowanych przez najlepszych badaczy problemów, a także wytrwałość w uzupełnianiu szczegółów dość skomplikowanych rozwiązań. Potwierdzeniem wysokiej jakości wyników rozprawy jest ich publikacja na prestiżowych konferencjach informatycznych SODA i ESA, a także specjalistycznej konferencji IPCO. Wprawdzie publikacje te są współautorskie, jednak taka jest natura badań we współczesnej informatyce.

Nie mam najmniejszych wątpliwości, że wyniki mgr. Bartosza Rybickiego spełniają ustawowe i zwyczajowe wymagania stawiane rozprawom doktorskim. Pewien niedosyt budzi jakość redakcyjna rozprawy. Mimo to, mając na uwadze wysoki poziom uzyskanych wyników, składam się do wnioskowania o wyróżnienie rozprawy.