

Profesor Lucjan Piela (em.)
Pracownia Chemii Kwantowej
Wydział Chemii
Uniwersytet Warszawski
Pasteura I
02-093 Warszawa

Warszawa, dn.30.VIII.2016 r.

Opinia o rozprawie doktorskiej p.mgr Anny Styrzcz zatytułowanej „Hybrydowe algorytmy optymalizacji globalnej – poszukiwanie niskoenergetycznych konformacji cząsteczek chemicznych”

Przedstawiona mi do oceny praca doktorska została wykonana w Zakładzie Chemii Teoretycznej im.Kazimierza Gumińskiego Wydziału Chemii Uniwersytetu Jagiellońskiego. Sam Zakład został dawno temu, można powiedzieć, „zaprogramowany” z pełną świadomością przez jego założyciela – prof.Kazimierza Gumińskiego. Liczyła się jakość osiągana talentem i ciężką pracą. Pani Anna Styrzcz jest w tym planie i nomenklaturze profesora Gumińskiego „doktorem V już rzutu”, jej promotorem jest p.prof.Jacek Korchowicz z tej samej szkoły naukowej. Myślę, że i niniejsza recenzja potwierdzi skuteczność dalekosiężnego planowania naukowego profesora Kazimierza Gumińskiego.

Temat rozprawy jasno przebija z tytułu. Może jedynie przymiotnik „hybrydowe” wymaga myślowego powiązania z „metodami czy obiektami mieszanymi” czyli zapowiada pewną niejednorodność podejścia. Waga opracowywanego zagadnienia jest bezdyskusyjna. Chemia w XX wieku wyrosła ze swojego ekscytującego dzieciństwa (atomy, wiązanie chemiczne, podstawowe prawa stereochemii) i dąży ku tworzeniu wielkich agregatów molekularnych, które wykonują jakąś funkcję. Aby być tu skutecznym, chemia musi pokonać trudność podstawową: panowania nad kształtem molekuł i ich syntezą tak, by molekuły złożyły się same, jak klocki Lego (rozpoznanie molekularne), w agregat o zadanych wcześniej, pożądanym właściwościach („programowanie molekularne” czy też tworzenie „materii poinformowanej”). Tu na każdym kroku przeszkodą są miriady możliwych konformacji, a bardzo często tylko *kilka* z nich decyduje o właściwościach fizykochemicznych układu. Teoretykowi nie pomaga w ich wyznaczeniu niemal nic... To dramat chemii i przedstawiona praca

atakuje właśnie to podstawowe i niezwykle trudne, nawet, powiedziałbym, dla chemii centralne zagadnienie, choć to ciągle tylko początki na trudnej drodze.

Znajdowanie ekstremum globalnego (będę o nim mówił jako minimum globalnym) w zasadzie nie mogło być przedmiotem badań przed nadejściem ery komputerów. Wtedy dopiero zwrócono uwagę, że zagadnienie optymalizacji, występujące przecież w zasadzie wszędzie w działalności człowieka, jest również immanentną cechą układów biologicznych i ekosystemów, które penetrują dostępne możliwości i rozwijają się w takim kierunku, który daje największe efekty („optymalizacja”).

To zaowocowało zaproponowaniem tzw. algorytmów genetycznych globalnej optymalizacji, które oparte są na założonej przez Darwina strategii ewolucji. Kluczowym krokiem jest sformułowanie funkcji celu (tu: energii układu) – jej minimum globalne w przestrzeni konfiguracji atomów ma być właśnie poszukiwane. Konfiguracje atomów są zakodowane przy użyciu pewnego zestawu parametrów, ich możliwe wartości definiują populację „osobników” (konformacji molekuł). Dalej, dokonywane są tzw. operacje genetyczne, zawsze na pewnej populacji osobników. I tak, można dokonać operacji selekcji wśród osobników jakiegoś rodzaju przekształcając jedną populację w inną. Kryterium (ale tylko jako preferencja przy losowaniu) jest wartość funkcji celu dla osobników – im niższa tym większe prawdopodobieństwo wylosowania. Jest to więc tendencja do koncentrowania się na znajdowaniu osobników w pobliżu osobnika najbardziej przystosowanego (z aktualnym minimum energii). Inną możliwą operacją na osobnikach jest mutacja (wylosowana zmiana jakiegoś parametru definiującego osobnika), a także krzyżowanie (również stochastycznie wybierana wymiana całej sekwencji parametrów między osobnikami tworząca nowych osobników). Obie te operacje mają za zadanie rozszerzenie zakresu poszukiwań w nadziei na znalezienie obszarów o niskiej energii czyli o „wysokim przystosowaniu”.

Selekcja, mutacja i krzyżowanie definiowane są losowo w pewnych protokołach o charakterze technicznym, same protokoły zależą od dość arbitralnych parametrów wybranych przez użytkownika, które mogą ulegać zmianie nawet w trakcie pojedynczej optymalizacji. Jak przekonała się o tym Doktorantka, nie istnieje uniwersalny zestaw parametrów dla algorytmu genetycznego, nawet ograniczając się do zagadnień molekularnych. Autorka stwierdza, że jest tak z powodu wielkich różnic w przebiegu funkcji celu dla różnych typów molekuł. To jest już „kuchenna strona”

optymalizacji - ważna, lecz łatwa do zaatakowania. Jednak, w kuchni jak to w kuchni, liczy się końcowy efekt (zadowolenie gości), a ten mistrzowie *haute cuisine* uzyskują na wiele różnych sposobów! Zaproponowanie sterowania wartościami tych parametrów za pomocą sieci neuronowej tak, aby maksymalizować sukces, czyli „uczenie sieci neuronowej rozsądnego działania”, jest oryginalnym i, moim zdaniem, najciekawszym wkładem Autorki. Udowodniono, że algorytmy związane z sieciami neuronowymi mają pewne archetypiczne cechy, które mogą być powiązane ze świadomością i generalizacjami dokonywanymi przez mózg. Algorytmy genetyczne i sieci neuronowe były oddzielnymi dziedzinami (choć oba podejścia oparte są przecież na idei optymalnego dostosowania), tu połączono je w sposób „hybrydowy” - stąd tytuł rozprawy, zawiadując automatycznie i inteligentnie algorytmem ewolucyjnym w celu szybszego otrzymania niskoenergetycznych konformacji molekuly. Dodać trzeba, że te konformacje są zawężone przez korzystanie z minimalizacji lokalnej (a więc operowanie pozycjami minimów energii) a to powoduje, że algorytm Autorki należy do tzw. algorytmów memetycznych.

Autorka rozprawy w sposób godny podziwu stara się stać na twardym gruncie matematyki. Mam tu akurat na myśli jej załączone do rozprawy piękne studium minimalizacji czterech z funkcji używanych przez matematyków do testowania procedur optymalizacji globalnej. Są to funkcje Ackleya, Rastrigina, Griewanka i Dixona-Price'a. Nie mają one bezpośrednich powiązań z chemią, a ich cechą wspólną jest olbrzymia liczba fałszywych basenów, których maksymalna głębokość zmienia się jednak w sposób regularny. Eksperymenty numeryczne Doktorantki wskazują na bardzo ważną rolę implementacji sieci neuronowej do procedury algorytmu genetycznego (to wkład Doktorantki), choć nie zawsze – czasem statyczny algorytm genetyczny jest skuteczniejszy. Te ostatnie przypadki – wcale częste w eksperymentach numerycznych Autorki - sam składałbym na karb niedoskonałości parametrów sieci neuronowej. Doktorantka pisze: „*Nie udało się natomiast ustalić optymalnej topologii samej sieci neuronowej. Dla każdej badanej funkcji optymalna sieć miała nieco inne parametry*”. Autorka stwierdziła ponadto, że szczególnie oporną dla globalnej optymalizacji metodą algorytmów genetycznych z siecią neuronową była funkcja Griewanka. Mam do tej funkcji sentyment szczególny, gdyż akurat w jej przypadku nasza metoda (metoda DEM - J.Kostrowicki, L.Piela, J.Opt.Theory and Appl. 69, 269-284 (1991)) pozwalała znaleźć minimum

globalne ...bez obliczeń. Choć nie oznacza to, oczywiście, gwarantowania jakiegokolwiek przewagi DEM w przykładach molekularnych, to sugeruję Doktorantce poeksperymentowanie w jej własnym algorytmie ze zmodyfikowaną przez DEM funkcją celu. Ta modyfikacja w jej programie byłaby bardzo prosta do wykonania.

Algorytmy Doktorantki wykazały jeszcze jedną ciekawą i, moim zdaniem, istotną i pozytywną cechę. Mianowicie, jej procedury nie zachowują tożsamości chemicznej optymalizowanych konformacji molekuł! Inaczej, szukane i znajdowane są struktury o innym układzie wiązań chemicznych (niż molekula startowa), czyli obniżanie energii biegnie pozwalając zrywać i tworzyć wiązania. To ważne, bo jest czasem dla chemika zaskakujące, gdy dowiaduje się, iż jego molekuly to zaledwie stan metastabilny. Z drugiej strony, ponieważ czasem wyłącznie takim stanem metastabilnym powinniśmy się zajmować, więc Doktorantka obmyśliła na to metody obrony w swojej procedurze optymalizacyjnej.

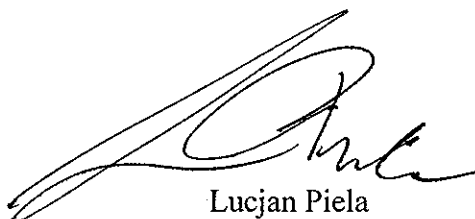
Molekuly, na których Autorka testowała swoje algorytmy nie są w obecnych czasach bardzo duże, choć te słowa padają w tej chwili z ust człowieka, który w pełni docenia „eksplozję konformacyjną” obezwładniającą jakiegokolwiek próby optymalizacji globalnej dla dużych molekuł. W zasadzie jednak, moim zdaniem, zarówno dla metenkefaliny, jak i ergotaminy, także ikozanu i dekanu, także dla dalej przytoczonych dwurdzeniowych kompleksów nitrozyłowych molibdenu czy nawet cyklodekstryny ciągle jeszcze byłaby możliwość użycia „brutalnej siły” przez enumerację możliwych konformacji i ich optymalizację (ew. wstępną) poprzez pole siłowe. Wierzę, że Autorka idzie szybciej w swoich algorytmach do tego celu. Według mojego przekonania dotychczasowe i spektakularne sukcesy optymalizacji globalnej w świecie wykazane nawet dla białek z 200 aminokwasami mają swoje źródło w innych niż algorytmy genetyczne podejściach. Należą do nich: rezygnacja z reprezentacji atomowej na korzyść pseudoatomów czyli wygładzanie hiperpowierzchni energii, wykorzystywanie banków struktur chemicznych takich jak Protein Data Bank przez (nieelegancką, przyznaję) konstrukcję na ich podstawie empirycznych potencjałów oddziaływania pseudoatom-pseudoatom i użycie metod Monte Carlo (w funkcji temperatury).

Podsumowując, rozprawa stoi na wysokim poziomie merytorycznym i technicznym. Jest dbałość o to, aby przedstawiać algorytmy w sposób matematycznie ścisły, czasem przez pomijanie konotacji chemicznych daje to wrażenie zbytnej

„suchości”. Widać, że Doktorantka jest biegła w wykorzystaniu gotowych programów zarówno dotyczących algorytmów genetycznych jak i sieci neuronowych. Jest także biegła w opisie trudnych zagadnień algorytmicznych. Opis zagadnień i z jednej i z drugiej dziedziny jest wysokiej jakości, choć brak mi trochę (zwłaszcza w pierwszym opisie) zacięcia dydaktycznego. Dla przyszłości Doktorantki moja uwaga może być istotna, bo prac doktorskich nie pisze się tylko dla bardzo wąskiego grona specjalistów, ale mogłyby one posłużyć do tworzenia pewnej generalnej kultury przedstawiania trudnych spraw.

Ogólna moja ocena rozprawy jest bardzo dobra, powiedziałbym nawet, że jest to rozprawa wyróżniająca się dobrym autorskim pomysłem (algorytm genetyczny sprzężony z siecią neuronową), bardzo starannym testowaniem tego pomysłu (oparciem się na precyzji bliższej informatykowi niż chemikowi), ale później zastosowaniem pomysłu do molekuł ważnych dla biologii i biochemii. Wrażenie robi także rzetelna analiza pokazująca sukcesy, ale także i porażki algorytmu. Praca jest więc rzeczywiście rozprawą i to rozprawą na wysokim poziomie.

Wniosuję o dopuszczenie Doktorantki do dalszych etapów przewodu doktorskiego.



Lucjan Piela