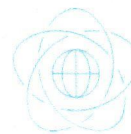


Wydział Matematyki i Nauk Informacyjnych Politechnika Warszawska

Koszykowa 75/507
00-662 Warsaw

prof. dr hab. inż. Przemysław Biecek
email: Przemyslaw.Biecek@pw.edu.pl



Warszawa, 27 stycznia 2025

Recenzja osiągnięcia naukowego „Uczenie zmiennych zbiorowych z symulacji atomistycznych” zgłoszonego przez dr. inż. Jakuba Rydzewskiego w procesie nadania stopnia doktora habilitowanego.

Pan dr inż. Jakub Rydzewski zgłosił jako osiągnięcie naukowe w procesie habilitacyjnym cykl siedmiu powiązanych tematycznie artykułów opublikowanych w czasopismach naukowych. Artykuły te we wniosku zostały oznaczone symbolami H1-H7.

Poniższa recenzja krótko charakteryzuje poszczególne artykuły a następnie podsumowuje ich wkład w rozwój dyscypliny naukowej nauki fizyczne. Artykuły te korzystają z metod i intuicji z obszaru uczenia maszynowego w rozwiązywaniu problemów z dyscypliny nauk fizycznych. Moje doświadczenie dotyczy głównie obszaru uczenia maszynowego i na tej części osiągnięcia skupię się w poniższej recenzji.

Praca H1 “Manifold Learning in Atomistic Simulations: A Conceptual Review” to praca przeglądowa omawiająca techniki konstrukcji niskowymiarowych rozmaitości rekonstruujących istotne charakterystyki oryginalnych wysoko wymiarowych opisów procesów fizycznych.

**Politechnika
Warszawska**

ul. Koszykowa 75
00-662 Warszawa

Wydział Matematyki i Nauk Informacyjnych Politechnika Warszawska

Koszykowa 75/507
00-662 Warsaw

prof. dr hab. inż. Przemysław Biecek
email: Przemyslaw.Biecek@pw.edu.pl



Konstrukcja niskomiarowych reprezentacji (ang. manifold learning) wspomaga proces próbkowania, który jest stosowany w przeprowadzaniu symulacji układów dynamicznych opisujących układy fizyczne (np. strukturę białka). Ta praca w spójny sposób opisuje techniki statystyczne stosowane dla redukcji wymiarowości, takie jak DMAP, TICA, SGOOP, SNE, UMAP, oraz zaproponowaną przez autora metodę MRSE, w odniesieniu do zadań związanych z charakteryzacją symulacji układów dynamicznych. Praca tworzy ramy uspołniające nomenklaturę i wprowadzającą pojęć do opisu problemów, które poruszane są w innych artykułach osiągnięcia naukowego.

Praca H2 "Multiscale Reweighted Stochastic Embedding: Deep Learning of Collective Variables for Enhanced Sampling" wprowadza nową metodę MRSE (Multiscale Reweighted Stochastic Embedding, wymienioną w pracy przeglądowej H1) oraz przedstawia przykłady jej zastosowań dla trzech układów modelowych. Metoda MRSE bazuje na bardzo popularnej metodzie redukcji wymiaru, jaką jest t-SNE (t-distributed stochastic neighbour embedding), ale operuje na parametryzowalnych transformacjach redukujących wymiarowość danych. Parametryzowalna transformacja jest uczona automatycznie z użyciem sieci neuronowych (automatycznie w przeciwieństwie do semiautomatycznego wyboru zmiennych zbiorowych - procesu angażującego wiedzę dziedzinową i intuicję eksperta dziedzinowego). W pracy przedstawione są trzy scenariusze zastosowania tej metody. Tabela 1 w tej pracy przedstawia architekturę sieci neuronowych zastosowanych do konstrukcji parametryzowalnej transformacji. Są one względnie płytke w stosunku do sieci dzisiaj stosowanych w uczeniu maszynowym w pracy nad złożonymi modalnościami takimi jak tekst czy obraz. Sam fakt, że stosowana transformacja jest uczona w postaci parametryzowalnej funkcji ma dużą przewagę w stosunku do algorytmu t-SNE, ponieważ skoro jest deterministyczna to można ją stosować w inferencji dla nowych danych.

Politechnika
Warszawska

ul. Koszykowa 75
00-662 Warszawa

Wydział Matematyki i Nauk Informatycznych Politechnika Warszawska

Koszykowa 75/507
00-662 Warsaw

prof. dr hab. inż. Przemysław Biecek
email: Przemyslaw.Biecek@pw.edu.pl



Proponowana metoda MRSE jest algorytmem z szeregiem usprawnień adaptujących intuicję z uczenia maszynowego do problemów w charakteryzacji procesów dynamicznych. Autorzy pokazują na przykładach jej użyteczność w rozwiązywaniu problemów w dyscyplinie nauki fizyczne.

Praca H3 "Reweighted Manifold Learning of Collective Variables from Enhanced Sampling Simulations" rozwiązuje problem skrzywionej (biased) estymacji zmiennych zbiorowych (collective variables). Proponowane rozwiązanie oparte jest o procedurę ważenia próbek, która ma solidne podstawy matematyczne, opisane w załączniku tej pracy. Zaproponowane diffusion reweighting można zastosować do treningu różnorodności a tym samym do identyfikacji zmiennych zbiorowych. Wyniki ważenia są dyskutowane na prostym układzie modelowym, który pokazuje skuteczność procesu ważenia. Następnie wynik tej metody jest ilustrowany na bardziej rzeczywistym przykładzie dla małej proteiny Chignolin. Zaproponowana w tym artykule metoda jest zaimplementowana we frameworku informatycznym do symulacji molekularnych rozwijanym przez konsorcjum PLUMED (czyli jest też częścią pracy H7 tego osiągnięcia naukowego).

Krótką pracę H4 "Selecting High-Dimensional Representations of Physical Systems by Reweighted Diffusion Maps" jest rozszerzeniem/kontynuacją pracy H3 i w pewnym sensie H1. Problem, który ta praca adresuje to wybór wysoko wymiarowej reprezentacji, na bazie której można skonstruować lepszą niskowymiarową reprezentację. Praca wprowadza prostą heurystykę wyboru zmiennych (w statystyce tego typu algorytmy czasem określa się przez step-up step-down) oraz przedstawia przykłady jej zastosowania do analizy danych modelowych i rzeczywistych.

Praca H5 "Spectral Map: Embedding Slow Kinetics in Collective Variables" proponuje nową metodę mapy spektralnej oparte o k najważniejszych zmiennych zbiorowych.

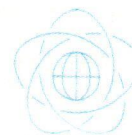
**Politechnika
Warszawska**

ul. Koszykowa 75
00-662 Warszawa

Wydział Matematyki i Nauk Informatycznych Politechnika Warszawska

Koszykowa 75/507
00-662 Warszawa

prof. dr hab. inż. Przemysław Biecek
email: Przemyslaw.Biecek@pw.edu.pl



Autor stosuje terminologię „szybkie” i „wolne” zmienne zbiorowe, która w klasycznym uczeniu maszynowym odpowiadałaby potencjałowi zmiennych/ważności zmiennych do opisu zachowania systemu. W proponowanej metodzie znajdują się dwie ciekawe propozycje, jedna to zastosowanie sieci neuronowych do uczenia się niskowymiarowej reprezentacji (widzieliśmy to już w H2 i jest to znaczące usprawnienie w stosunku do konstrukcji zmiennych zbiorowych przez liniową agregację), a druga to wykorzystanie „przerwy spektralnej” do rozdzielenia zmiennych „szybkich” od „wolnych”. Wyznaczanie przerwy spektralnej pomiędzy wartościami spektralnymi dla Markowskiej macierzy przejścia przypomina algorytm „osypiska” lub „metodę łokcia” stosowaną w klasycznym problemie wyznaczenia optymalnej liczby skupisk w uczeniu maszynowym. Ta koncepcja jest jednak przeniesiona na nowy problem konstrukcji zmiennych zbiorowych.

Praca H6 “Learning Markovian Dynamics with Spectral Maps” jest rozszerzeniem pracy H5 adaptującym je do analizy systemów w długiej osi czasu. Dla takich skal czasowych, kluczowy jest właściwy wybór zmiennych grupowych by opis z użyciem dynamiki Markowskiej był wierny/poprawny. Autor ilustruje jak w tym celu wykorzystać adaptacyjne dobierany zestaw zmiennych konstruowanych z użyciem algorytmu „przerwy spektralnej” przedstawionego w artykule H5.

W pracach H1-H6 autor ma wiodący wkład w prezentowane osiągnięcie. Przedstawione w nich metody analityczne są zaimplementowane w oprogramowaniu rozwijanym przez konsorcjum PLUMED. Praca H7 “Promoting Transparency and Reproducibility in Enhanced Molecular Simulations” opisuje zbiór narzędzi informatycznych i analitycznych pozwalających na analizę symulacji molekularnych. Udostępnianie na otwartych licencjach opracowanego oprogramowania to oczywiście bardzo dobra praktyka.

Politechnika
Warszawska

ul. Koszykowa 75
00-662 Warszawa



Powyższe podsumowanie było jednoznacznie pozytywnym omówieniem potencjału zaproponowanych metod. Jedyną krytyczną uwagę, którą mam do zaprezentowanych badań, to niedosyt związany z interpretowalnością proponowanych rozwiązań. Pomimo iż, podkreślenie wagi interpretowalności pojawia się kilkakrotnie w opisie osiągnięcia, to autor nie zastosował ani nie zaprezentował technik popularnych o obszarze eXplainable Artificial Intelligence (XAI), do analizy sieci neuronowych, wykorzystanych do uczenia się parametrycznej niskowymiarowej reprezentacji danych. Skoro już wytrenowane są sieci konstruujące niskowymiarową reprezentację, to analiza XAI takich sieci mogłaby doprowadzić do ciekawych obserwacji i lepszego zrozumienia zjawisk fizycznych. Zachęcam autora do rozważenia zastosowania takich technik jak SmoothGrad, GradCam, SaliencyMaps, IntegratedGradients itp do analizy tej transformacji.

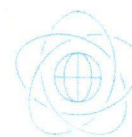
We wniosku autor przedstawił również szereg innych publikacji, oraz aktywności projektowych, grantowych, związanych z popularyzacją badań, aktywnym uczestnictwem w procesie obiegu nauki. Autor osiągnięcia aktywnie współpracuje z dobrymi międzynarodowymi zespołami badawczymi.

Podsumowując. Prace H1-H7 są spójne tematycznie i przedstawiają szereg ciekawych nowych metod analizy danych (w tym Multiscale Reweighted Stochastic Embedding oraz Spectral maps), wraz pracą przeglądową umieszczającą te metody na tle innych podejść do redukcji wymiaru reprezentacji opisu symulacji procesów dynamicznych (H1) oraz pracę podsumowującą implementacje opracowanych rozwiązań (H7). Zaproponowane rozwiązania czerpią z wielu rozwiązań z obszaru uczenia maszynowego, są jednak nowymi metodami adaptującymi te intuicje do nowych potrzeb w analizie zjawisk fizycznych. Zaproponowane metody są zilustrowane przykładami zastosowań symulacji układów dynamicznych na poziomie molekularnym.

Wydział Matematyki i Nauk Informatycznych Politechnika Warszawska

Koszykowa 75/507
00-662 Warszawa

prof. dr hab. inż. Przemysław Biecek
email: Przemyslaw.Biecek@pw.edu.pl



W mojej opinii, te nowe metody do niskowymiarowej charakterystyki złożonych układów stanowią istotny wkład w rozwój dyscypliny nauk fizycznych, a zaprezentowane przykłady jedynie potwierdzają, że mają one szeroki wachlarz zastosowań do analizy zjawisk fizycznych. Prace te wpisują się w coraz bardziej obiecujący nurt badań interdyscyplinarnych wykorzystujących istniejące oraz tworzenie nowych technik uczenia maszynowego (czy szerzej sztucznej inteligencji) do modelowania zjawisk fizycznych i chemicznych. Autor nie tylko wprowadza nowe metody ale też ilustruje ich działanie oraz zalety na rzeczywistych przykładach. Wyniki swoje publikuje w renomowanych czasopismach międzynarodowych.

Biorąc pod uwagę powyższe, uznaję, że osiągnięcie naukowe dr inż. Jakuba Rydzewskiego, na które składa się cykl siedmiu artykułów naukowych, spełnia z naddatkiem wymagania Ustawy z dnia 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce, w szczególności Artykułu 219 dotyczące nadania stopnia doktora habilitowanego.

Przemysław Biecek

Politechnika
Warszawska

ul. Koszykowa 75
00-662 Warszawa