



dr hab. inż. Paweł Morawiecki, prof. IPI PAN

Warszawa, 20.09.2023

Recenzja rozprawy doktorskiej „Interpretable Deep Learning with Prototypical Parts for Supervised and Weakly-Supervised Learning” autorstwa Dawida Rymarczyka

Tematyka rozprawy

Tematyka rozprawy dotyczy algorytmów głębokiego uczenia, które oprócz predykcji zwracają również interpretowalną informację. W ten sposób uwiarygadniane jest działanie całej sieci neuronowej. Zagadnienie interpretowalności czy wyjaśnialności jest niezwykle ważne w obszarach takich jak medycyna czy bezpieczeństwo, gdzie błędna, mało zrozumiała decyzja modelu może nieść za sobą poważne konsekwencje. Autor rozprawy skupił się na metodach „samowyjaśnialnych”, które nie angażują dodatkowego modelu już po dokonanej predykcji.

Charakterystyka rozprawy

Głównym celem rozprawy postawionym przez Doktoranta jest

„zbudowanie bardziej transparentnych i wiarygodnych modeli uczenia maszynowego bazujących na prototypach i mechanizmie uwagi”.

Postawiony cel/hipoteza weryfikowana jest przede wszystkim w odniesieniu do modeli wizyjnych w kontekście zagadnień klasyfikacyjnych, w tym również medycznych zbiorów danych.

Praca składa się z cyklu 5 publikacji. Przewodnik po publikacjach zawiera krótkie omówienie podstawowych terminów i pojęć oraz charakterystykę każdego z artykułów. W lekturze pomaga omówienie wzajemnych relacji między pracami i przedstawienie motywacji stojących za poszczególnymi rozwiązaniami. Dominującym motywem pojawiającym się w pracach to wykorzystanie prototypów (charakterystycznych cech czy wzorców dla danej klasy).

Cykl publikacji

Poniżej przywołuję publikacje wchodzące do przedstawionego cyklu wraz z krótkim ich podsumowaniem.

[I] Dawid Rymarczyk, Łukasz Struski, Jacek Tabor, Bartosz Zielinski. “Protopshare: Prototypical parts sharing for similarity discovery in interpretable image classification.” ACM Conference on Knowledge Discovery & Data Mining (KDD), 2021, 1420-1430

W tej pracy wprowadzono model do interpretacji klasyfikacji obrazów oparty na prototypowych częściach. Autorzy zaproponowali nową miarę podobieństwa zależną od danych, który pozwala na identyfikację semantycznej spójności między pojęciami wizualnymi, oraz operację scalania i przycinania (ang. merge and prune), która umożliwia współdzielenie prototypowych części. Uzyskane wyniki są lepsze od pierwowzoru czyli architektury ProtoPNet.

[II] Dawid Rymarczyk, Łukasz Struski, Michał Górszczak, Koryna Lewandowska, Jacek Tabor, Bartosz Zielinski. “Interpretable image classification with differentiable prototypes assignment.” European Conference on Computer Vision (ECCV), 2022, 351-368

W artykule wprowadzono nową architekturę ProtoPool, która zapewnia lepsze współdzielenie i wykorzystanie części prototypowych niż wcześniejsze rozwiązanie ProtoPShare. Ponadto wprowadzono nową metrykę podobieństwa (Focal Similarity), która daje lepszą interpretowalność. Badania użytkowników potwierdziły zalety nowego rozwiązania.

[III] Dawid Rymarczyk, Daniel Dobrowolski, Tomasz Danel. “ProGReST: Prototypical Graph Regression Soft Trees for Molecular Property Prediction.” SIAM Conference on Data Mining (SDM), 2023, 379-387

Praca skupia się na zadaniu przewidywania własności cząsteczek. Zaproponowane podejście dostarcza wglądu w model regresji jako interpretowalnej podstruktury grafu i struktury drzewa, co pozwala nam lepiej zrozumieć mechanizmy działania. Warto też podkreślić, że wprowadzona technika „proxy projection” umożliwia trening dużych modeli w akceptowalnym czasie.

[IV] Dawid Rymarczyk, Adriana Borowa, Jacek Tabor, Bartosz Zielinski. “Kernel self-attention for weakly-supervised image classification using deep multiple instance learning.” IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2021, 1721-1730

Praca łączy ze sobą mechanizm „self-attention” z poolingiem typu MIL (ang. multiple instance learning). Osiągnane wyniki są na poziomie state-of-the-art z jednoczesnym zapewnieniem interpretowalności. Rozwiązanie zostało sprawdzone na medycznych zbiorach danych dotyczących nowotworów piersi i jelita.

[V] Dawid Rymarczyk, Adam Pardyl, Jarosław Kraus, Aneta Kaczynska, Marek Skomorowski, Bartosz Zielinski. “ProtoMIL: Multiple Instance Learning with Prototypical Parts for Whole-Slide Image Classification.” European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), 2022

Autorzy pracy stosują wnioskowaniem bazujące na prototypach do problemu MIL (multiple instance learning). Zaprojektowana architektura pozwala nie tylko zinterpretować konkretny przykład, ale też scharakteryzować poszczególne klasy na podstawie uzyskanych prototypów. Metoda została sprawdzona na dużych zbiorach danych medycznych pochodzących z biopsji.

Wyżej przytoczone prace zostały przedstawione na bardzo dobrych konferencjach (ranga A lub A*).

Uwagi

Przedłożony cykl prac ma wysoką wartość merytoryczną, zaproponowane rozwiązania układają się w spójną całość. Zagadnienie interpretowalności ma bardzo duże znaczenie w momencie rzeczywistej implementacji wytrenowanych modeli. W szczególności w

sytuacjach, gdzie mamy do czynienia z kooperacją na linii człowiek-algorytm, takich jak diagnoza medyczna czy ocena kompetencji potencjalnego pracownika.

W przewodniku po załączonych pracach zabrakło mi dokładniejszego opisu „Prototype Layer”, która jest główną innowacją architektury ProtoPNet. W przypadku mechanizmu uwagi taki opis i definicje zostały wprowadzone (paragraf 3.2.2, strona 34).

Konkluzja

Cel badawczy postawiony przez Doktoranta został zrealizowany. W pracach zostały rozwiązane nietrywialne problemy naukowo-badawcze z silną motywacją praktyczną.

Uważam, że złożona rozprawa mgr Dawida Rymarczyka spełnia wymagania ustawowe i zwyczajowe stawiane pracom doktorskim i może stanowić podstawę nadania stopnia doktora. Ponadto stwierdzam, że przedłożony cykl publikacji wykracza zdecydowanie poza zwykłe wymagania stawiane rozprawom doktorskim w informatyce technicznej. Podjęta tematyka jest poznawczo bardzo ciekawa i społecznie ważna, a uzyskane rezultaty opublikowane na bardzo dobrych konferencjach. Dlatego też wnioskuję o wyróżnienie rozprawy.

Prof. Halimeda