

Abstrakt

Modele uczenia głębokiego (oparte na głębokich sieciach neuronowych) są postrzegane jako tzw. czarne skrzynki, które są trudne do interpretacji ze względu na dużą liczbę trenowalnych parametrów. Jednak z uwagi na wymogi prawne oraz oczekiwania interesariuszy, opracowano różne techniki mające na celu lepsze wyjaśnienie sposobu analizowania danych i podejmowania decyzji przez modele. Informacja o tym jak działają różne komponenty sieci neuronowej jest ważna dla osób tworzących modele. To jest celem tzw. czerwonej wyjaśnialnej sztucznej inteligencji (ang. Explainable AI, XAI), która stawia na pierwszym miejscu wyjaśnienia służące technicznej walidacji modelu, w przeciwieństwie do niebieskiej XAI, koncentrującej się na dopasowywaniu wyjaśnień do sposobu myślenia ludzi i ich wartości.

Podczas podejmowania decyzji istotne jest, aby dysponować możliwie jak najbardziej kompletnymi informacjami. Często obejmuje to informacje kontekstowe, które pozwalają na holistyczną ocenę sytuacji i w efekcie bardziej świadome decyzje. Zasada ta powinna również odnosić się do modeli uczenia głębokiego. Informacje kontekstowe mają szczególne znaczenie w przetwarzaniu danych pochodzących z systemów monitoringu ulicznego, samochodów autonomicznych, robotów, opieki zdrowotnej, rolnictwa itp.

Istnieją różne rodzaje informacji kontekstowej, w tym przestrzenna i temporalna, które są głównym przedmiotem analiz w niniejszej rozprawie. Postrzeganie kontekstu przestrzennego przez modele oparte na sieciach neuronowych jest badane na przykładach cyfrowej histopatologii oraz wizualnych testów IQ. Z kolei kontekst temporalny analizowany jest w modelu służącym do klasyfikacji tweetów w trybie ciągłym. Prowadzone badania są zgodne z założeniami czerwonej XAI.

Ponadto pokazano, że popularne techniki XAI nie uwzględniają relacji przestrzennych pomiędzy obiektami w obrazach wejściowych, ograniczając się jedynie do analizy roli pojedynczych obiektów w procesie podejmowania decyzji przez model. Oznacza to, że modele głębokie nie mogą być właściwie ewaluowane, gdy istotna jest informacja kontekstowa. Brak dedykowanych metod do analizy takich relacji został wskazany jako luka badawcza, którą niniejsza praca doktorska podejmuje.

Opracowano szereg rozwiązań umożliwiających badanie roli kontekstu przestrzennego o różnym rozmiarze (od lokalnego do globalnego) w sieciach neuronowych stosowanych do klasyfikacji tkanek. Aby zbadać rolę kontekstu lokalnego w modelach konwolucyjnych i opartych na transformerach do klasyfikacji łątek (ang. *patches*) przedstawiających tkankę, zaproponowano metodę opartą na zasłonięciu kontekstu. Badano, czy modele są wrażliwe na brak informacji kontekstowej w postaci tkanki otaczającej analizowaną łątkę. Ponadto oceniono rolę kontekstu semiglobalnego i globalnego (kompozycyjności obrazu i relacji sąsiedztwa dalekiego zasięgu) w modelach opartych na wielu instancjach (ang. *Multiple Instance Learning*), które są często wykorzystywane do analizy obrazów wysokiej rozdzielczości, takich jak obraz całej tkanki (ang. *Whole Slide Images*) w histopatologii cyfrowej. W tym celu opracowano trzy zestawy danych (*SyntheticSMIL*), protokół

ewaluacyjny (*ReasonSMIL*) oraz metodę *Deep Spatial Context*. Metoda *Deep Spatial Context* jest matematycznie ugruntowanym podejściem do ilościowej oceny rozumowania w oparciu o informację przestrzenną. Metoda ta razem z tą opartą na zasłanianiu kontekstu poszerzają arsenal technik w ramach “przestrzennej wyjaśnialnej sztucznej inteligencji”. Przeprowadzono również analizę porównawczą skuteczności modeli trenowanych na zdjęciach całych tkanek a tych trenowanych na podzbiorze łatek (regionów) z każdego oryginalnego zdjęcia. Wyniki eksperymentalne wskazują, że dodatkowy wysiłek włożony w wybór kluczowych regionów przez lekarza nie przekłada się na lepszą skuteczność modeli w porównaniu z wariantem, gdy użyte są surowe dane z całymi tkankami. Zaobserwowano, że podczas testowania modeli na surowych danych w ramach inter-ewaluacji, najlepszą skuteczność, spośród modeli trenowanych na podzbiorach regionów, osiągnął model trenowany na losowo wybranych regionach. Tym samym był to lepszy model niż te trenowane na regionach wybranych przez lekarza oraz regionach wyselekcjonowanych na podstawie najwyższych wartości uwagi z modelu wytrenowanego na surowych danych.

Ponadto, kontekst przestrzenny został uwzględniony w projektowaniu algorytmu *Duel-IQ* do rozwiązywania wizualnych testów IQ, które wymagają zrozumienia relacji przestrzennych.

Rola kontekstu temporalnego była eksplorowana w ramach zaproponowanej metody *LIMES*, w której kontekst temporalny jest wykorzystany w mechanizmie adaptacji do zmiennego rozkładu klas w danych.

Podsumowując, w ramach rozprawy doktorskiej pokazano, że informacja kontekstowa (zarówno przestrzenna, jak i temporalna) odgrywa kluczową rolę w klasyfikacji realizowanej przez głębokie sieci neuronowe. Jest to obserwowane przy różnych rozmiarach kontekstu (od lokalnego do globalnego) i w różnych obszarach.

Słowa kluczowe: kontekst przestrzenny, cyfrowa histopatologia, modele oparte na wielu instancjach, kontekst przestrzenny, wyjaśnialna sztuczna inteligencja.