Częstochowa, dn. 11 października 2024 r.

prof. dr hab. inż. Rafał Scherer
Katedra Sztucznej Inteligencji
Wydział Informatyki i Sztucznej Inteligencji
Politechnika Częstochowska
al. Armii Krajowej 36
42-200 Częstochowa

Recenzja

rozprawy doktorskiej mgra Macieja Żelaszczyka, pt.: Deep Representation Learning in Varied Settings.

Niniejszą recenzję opracowano na wniosek Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej. Promotorem jest prof. dr hab. inż. Jacek Mańdziuk.

1. Charakterystyka tematu, celu i tezy badawczej rozprawy

Uczenie reprezentacji, zwłaszcza międzymodalne, ma kluczowe znaczenie, ponieważ pozwala na integrację informacji pochodzących z różnych źródeł, takich jak dźwięk, obraz, tekst czy wideo. Dzięki temu możliwe jest zrozumienie i przetwarzanie danych w sposób bardziej zbliżony do ludzkiej percepcji, gdzie informacje z różnych zmysłów są łączone, aby tworzyć pełniejszy obraz rzeczywistości. W uczeniu między modalnym istotną korzyścią jest zdolność modelu do przenoszenia wiedzy między różnymi typami danych, co ma ogromne znaczenie w praktyce. Reprezentacje między modalne może pomóc w efektywniejszym przetwarzaniu danych w sytuacjach, gdy dostęp do jednej z modalności jest ograniczony. Równie istotne znaczenie ma przeciwdziałanie atakom które celowo wprowadzają drobne zmiany do danych wejściowych (np. obrazów, tekstu), aby zmylić model uczenia maszynowego i sprawić, że poda błędne wyniki. Praca adresuje wybrane problemy występujące w tworzeniu reprezentacji z danych, w tym w uczeniu wielozadaniowym.

2. Zawartość rozprawy

Recenzowana praca mgra Macieja Żelaszczyka składa się z siedmiu rozdziałów, bibliografii spisu algorytmów i rysunków. Dokument liczy 160 stron.

Pierwszy rozdział jest wprowadzeniem do tematyki: omawia uczenie maszynowe jako uczenie reprezentacji, oraz podaje warunki, przy których reprezentacja będzie użyteczna. Podana jest ogólnie brzmiąca teza, stwierdzająca, że uczenie reprezentacji jest w stanie rozwiązywać odmienne od siebie zadania w zróżnicowanych dziedzinach. Na koniec omówiono zawartość rozprawy wraz z przypisaniem publikacji do poszczególnych zagadnień badawczych.

Rozdział 2 wprowadza czytelnika w zagadnienia związane z rozprawą. Podane są dwie definicje modeli generatywnych oraz podstawowe modele: autoenkoder wariacyjny i GAN. Następnie omówiono ataki adwersarialne polegające na celowym dobraniu wartości wejściowej modelu w celu uzyskania niespójnego wyniku z ogólna reprezentacją nauczoną przez model. Na koniec omówiono koncepcję osadzeń dla danych tekstowych i numerycznych.

Rozdział 3 proponuje metodę obrony przed atakami adwersarialnymi, w której dane wyjściowe przetwarzane są poprzez mieszaninę generatorów, aby odzyskać poprawną klasę sprzed ataku. Zestaw kanonicznych obrazów oraz zestaw ataków adwersarialnych zostały użyte do stworzenia wersji obrazów kanonicznych z zakłóceniami adwersarialnymi. Obrazy te są następnie przetwarzane przez komitet generatorów, które rywalizują o aktualizacje wag, próbując odwrócić transformacje adwersarialne. Zestaw samych generatorów jest trenowany w sposób adwersarialny. System był uczony bez nadzoru na wielu atakach adwersarialnych jednocześnie, a po nauczeniu potrafił odzyskać etykiety klas z nigdy wcześniej nieznanych próbek, nie mając dostępu do rzeczywistych etykiet ani informacji o tym, jaki konkretny atak został zastosowany. Eksperymenty pokazały, że zaproponowana metoda obrony wieloatakowej jest konkurencyjna względem obron zaprojektowanych przeciwko pojedynczym atakom adwersarialnym.

Rozdział 4 dotyczył tworzenia reprezentacji w zadaniach generowaniu obrazów z dźwięku co jest zadaniem bardzo trudnym z powodu czasami nieoczywistych zależności pomiędzy treściami audio a wizualnymi oraz inną specyfiką sygnałów czasowych i statycznych dwuwymiarowych. Autor stworzył dwa zbiory danych typu audio-obraz jako rozszerzenia istniejących źródeł audio i obrazów, gdzie wytworzone zbiory danych reprezentują odpowiednio mapowania audio na obraz w schemacie jeden-do-wielu i jeden-do-jednego. Obie architektury były uczone na powyższych dwóch zbiorach danych i wykazują zdolność do reprezentowania wejściowych danych audio jako cech audiowizualnych, czyli części reprezentacji dźwiękowej istotnej dla generowania obrazów. Używając tych reprezentacji oba zaproponowane modele potrafiły generować obrazy o wysokim stopniu podobieństwa wizualnego do rzeczywistych obrazów ze zbiorów danych. Eksperymenty pokazały również, że dla architektury trenowanej w sposób adwersarialny poziom zgodności między rzeczywistymi a generowanymi obrazami może być regulowany przez odpowiednie ważenie strat rekonstrukcji w ogólnej funkcji strat.

W rozdziale 5 zaproponowany jest interpretowalny system predykcyjny, w którym informacje są współdzielone między różnymi kontekstami predykcyjnymi w uczeniu wielozadaniowym. Rodzaje i wymiary danych wejściowych i wyjściowych mogą się różnić dla różnych zadań. Realizuje się to poprzez tworzenie osadzeń (embeddings) zmiennych wejściowych i wyjściowych, które istnieją w tej samej przestrzeni. Ostateczne osadzenia wejściowe powstają za pomocą mechanizmu uwagi, gdzie łączony jest wspólny zestaw osadzeń. Wspólne osadzenia są ponownie wykorzystywane między różnymi zadaniami predykcyjnymi. Wszystkie używane osadzenia, w tym wspólny zestaw, są traktowane jako parametry modelu i podlegają aktualizacjom podczas procesu uczenia.

W rozdziale 6 Autor zaproponował metodę ograniczania reprezentacji modelu predykcyjnego poprzez naukę zależności między komponentami w miarę postępu uczenia. Te wyuczone zależności są włączane do procedury uczenia kontradyktoryjnego (adwersarialnego), w której główny model optymalizuje swoją zdolność predykcyjną, jednocześnie próbując zapobiec możliwości przewidywania komponentów reprezentacji na podstawie innych komponentów. Realizuje się to poprzez trenowanie klasyfikatora oraz zestawu predyktorów, które mają przeciwstawne cele. Klasyfikator stara się zmniejszyć wartość swojej funkcji strat w zadaniu klasyfikacji, jednocześnie maksymalizując stratę zestawu predyktorów. Z kolei zestaw predyktorów minimalizuje swoją stratę przy przewidywaniu komponentu reprezentacji na podstawie innych komponentów. Eksperymenty pokazały, że wybór silnie skorelowanych reprezentacji dla predyktorów pozwala im uniknąć nadmiernego dopasowania do danych. Zaprezentowana metoda treningu kontradyktoryjnego działa na poziomie porównywalnym ze standardowym treningiem, jednocześnie znacząco zmieniając rozkład średnich komponentów reprezentacji w zbiorze testowym. Nie prowadzi jednak do dekoleracji reprezentacji. Silnie skorelowane komponenty są stosunkowo skoncentrowane i trwałe podczas treningu.

Rozdział 7 podsumowuje rozprawę oraz proponuje przyszłe kierunki badań.

3. Ocena rozprawy

W ramach rozprawy doktorskiej Pan mgr Maciej Żelaszczyk zaproponował zestaw oryginalnych rozwiązań związanych z tworzeniem reprezentacji z danych realizując cel pracy. Tematyka pracy jest bardzo aktualna i potrzebna, oryginalny dorobek autora polega na stworzeniu:

- metody obrony przed atakami na system uczenia maszynowego za pomocą mieszaniny generatorów,
- systemu generowania międzymodalnego treści wizualnych z danych audio.
- metody uczenia wielozadaniowego z dzielonymi zmiennymi osadzeniami,
- metod ograniczających tworzone reprezentacje,
- autorskich zbiorów danych.

Rozprawa doktorska uwidacznia wysoką ogólną wiedzę teoretyczną i praktyczną oraz umiejętność samodzielnego prowadzenia pracy naukowej mgra Macieja Żelaszczyka. Pan Żelaszczyk opracował wprowadzenie do tematyki i przegląd literatury na temat poszczególnych obszarów swoich badań oraz zadbał o popularyzację wyników swoich badań publikując liczne prace naukowe w materiałach uznanych konferencji oraz w czasopismach (m.in. 200 pkt).

Rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego. Zaproponowane metody mają duże znaczenie dla nauk technicznych oraz przemysłu, zarówno teoretyczne, jak i aplikacyjne. Praca napisana jest w języku angielskim na wyjątkowo dobrym poziomie.

4. Wnioski końcowe recenzji

Podsumowując recenzję stwierdzam, że Pan mgr Maciej Żelaszczyk, pt.: w rozprawie doktorskiej „Deep Representation Learning in Varied Settings" zrealizował cel rozprawy. Zaprezentowane rezultaty stanowią oryginalny wkład Autora w rozwój dyscypliny Informatyka Techniczna i Telekomunikacja. Pan Maciej Żelaszczyk wykazał się umiejętnością samodzielnej pracy badawczej, znajomością literatury światowej i wiedzą w zakresie uczenia maszynowego. Recenzowana praca spełnia wymagania ustawy o tytule i stopniach naukowych w dyscyplinie naukowej Informatyka Techniczna i Telekomunikacja. Wnoszę o jej przyjęcie i dopuszczenie do dalszych etapów postępowania doktorskiego.