Marcin Plata

Zastosowanie zaawansowanych metod sztucznej inteligencji do wybranych problemów bezpieczeństwa informacji

Rozprawa doktorska

Promotor:
prof. dr hab. inż. Marek Klonowski

Promotor pomocniczy:
dr inż. Piotr Syga

Instytut Podstaw Informatyki
Polskiej Akademii Nauk

Warszawa 2023

Streszczenie

W niniejszej rozprawie przedstawiamy metody zaliczane do szeroko pojętej sztucznej inteligencji, które z sukcesem aplikujemy do wybranych problemów z obszaru bezpieczeństwa informacji. Dotykamy dwóch głównych zagadnień – znakowania wodnego obrazów oraz biometrii. W przypadku biometrii pochylamy się nad trzema różnymi tematami – standardowej autoryzacji, odporności na atak podszywania się oraz ochrony prywatności. Każdy problem jest przedstawiony bazując na innych cechach biometrycznych, tj. odpowiednio konturze dłoni, głosie, mimice twarzy.

Trudność w tworzeniu rozwiązań do znakowania wodnego wynika głównie z konieczności uodpornienia opracowywanych metod na ataki oraz algorytmy kompresji, takie jak JPEG. Sam dobór ataków może znacząco wpłynąć na architekturę rozwiązania (np. operacja przycinania krawędzi obrazu zaburza jego strukturę przestrzenną, znacząco utrudniając odczytanie znaku wodnego, przez co jest często pomijana w pracach badawczych). W rozprawie prezentujemy nowe podejście oparte na trzech sieciach neuronowych – enkoder, dekoder oraz dyskryminator, które rozszerzamy o autorską metodę rozprzestrzeniania przestrzennego znaku wodnego, która wyraźnie zwiększa odporność na część ataków oraz kompresję JPEG w stosunku do ostatnich rozwiązań, a przy tym utrzymuje wysoki poziom odporności na atak przycinania. Następnie dodajemy do enkodera dodatkowy komponent, który przetwarza znak wodny niezależnie od obrazu wejściowego, tym samym uzyskując wyższą odporność oraz jego jakość, nie zwiększając rzeczywistego czasu nakładania znaku wodnego. Następnie proponujemy podział ataków na grupy, zależnie od operacji przetwarzania obrazu i pokazujemy, że w celu osiągnięcia wysokiej ogólnej odporności wymagane jest uwzględnienie ataków ze wszystkich grup. Zauważamy również fakt, że w rzeczywistych warunkach, treści oznaczone konkretną metodą znakowania wodnego występują stosunkowo rzadko, a tym samym istnieje zwiększone ryzyko błędnego oczytania znaku wodnego, prowadzące do fałszywego oskarżenia o piractwo. Ten problem jest przez nas adresowany poprzez wykorzystanie dyskrymina-

tora do wstępnego określenia istnienia znaku wodnego, następnie pokazujemy skuteczność naszego podejścia na podstawie opracowanych metryk.

Obecnie rozwiązania biometryczne są wykorzystywane na szeroką skalę, również jako metody autoryzacji do systemów wymagających szczególnych środków ostrożności, np. aplikacje bankowe. W rozprawie podejmujemy się tematu ochrony prywatności dla metod biometrycznych na przykładzie cech twarzy. W naszym rozwiązaniu staramy się ukryć typowe statyczne cechy i przetwarzać jedynie przesunięcia punktów charakterystycznych (mimikę). Takie podejście znacząco ogranicza możliwości rekonstrukcji twarzy z punktów charakterystycznych, np. w przypadku wycieku bazy danych. W celu zwiększenia prywatności dodatkowo normalizujemy cechy oraz proponujemy podejście, w którym odrzucamy najbardziej skorelowane cechy mimiki twarzy z cechami statycznymi. Nasze rozwiązanie zostało oparte na kilku klasycznych metodach uczenia maszynowego, takich jak maszyna wektorów nośnych czy drzewa losowe. Uzyskane wyniki pozwalają uznać naszą metodę jako tzw. biometrię słabą, która sprawdzi się również jako metoda wspomagająca autoryzację. Zastosowanie podejścia, w którym użytkownik musi wypowiedzieć określoną frazę w celu autoryzacji, pozwala rozważać nasze rozwiązanie jako biometrię usuwalną.

Metody do detekcji ataku spoofingowego są kluczowymi elementami systemów biometrycznych. Jednym z najbardziej skutecznych, a zarazem najtrudniejszym do wykrycia, atakiem jest powtórne odtworzeniowe nagranego głosu. Podejmujemy się stworzenia metody, która jest w stanie wykryć tego typu atak. Ze względu na specyfikę metody wykorzystujemy małe sieci neuronowe – LCNN oraz bayesowską sieć neuronową – umożliwiające szybką detekcję na podstawie kilkusekundowej próbki dźwiękowej. W celu wytrenowania modeli wykorzystujemy autorską metodę walidacji krzyżowej, uwzględniającą podział na ataki. Dodatkowo stosujemy szereg technik regularyzacyjnych, które umożliwiają wyuczenie rozwiązania wykrywającego również ataki spoza zbioru treningowego.

W przypadku metody do autoryzacji biometrycznej bazującej na konturze dłoni, pokazujemy, że wykorzystanie klasycznych metod przetwarzania obrazu jest wystarczające do stworzenia wydajnego obliczeniowo rozwiązania o wysokiej skuteczności. Nasz potok przetwarzania składa się z kilku szybkich algorytmów, takich jak rozmycie gaussowskie, liczenie otoczki wypukłej, wpisanie trójkąta o największym obwodzie w kontur. Takie podejście pozwala na wykorzystanie standardowego i taniego urządzenia, które nie wymaga GPU. Nasza metoda przyjmuje próbki pobrane skanerem biurowym, co dodatkowo pozwala na redukcję kosztów wdrożenia systemu.