

Karol ANTCZAK

Wojskowa Akademia Techniczna, Wydział Cybernetyki
ul. Gen. S. Kaliskiego 2, 00-908 Warszawa
E-mail: karol.antczak@wat.edu.pl

Uczenie głębokie w diagnostyce medycznej

1 Wstęp

Medycyna u swoich początków uważana była za sztukę. Przez wiele lat lekarze uważani byli uważani, w mniejszym lub większym stopniu, za posiadaczy swego rodzaju wiedzy tajemnej, którzy w niewytłumaczalny sposób potrafili najpierw rozpoznać chorobę, by później przeprowadzić jej skuteczne leczenie. Mimo że był to proces do pewnego stopnia empiryczny, opierał się on przede wszystkim na ukrytych, nieskodyfikowanych czynnikach, takich jak wiedza i przekonania własne lekarzy, znajomość reakcji osobniczej (idiosynkrazja) lub wiedza na temat skuteczności stosowanych technologii.

Przełom w empirycznym pojmowaniu medycyny przyniosło, upowszechnione w latach osiemdziesiątych, hasło „medycyny opartej na faktach” (ang. *Evidence Based Medicine*). Podejście to definiuje się jako „sumienne, jawne i rozsądne korzystanie z najlepszych dostępnych naukowo dowodów w podejmowaniu decyzji o leczeniu pacjentów” [1]. Medycyna oparta na faktach dużą rolę przykładu do korzystania z wyników badań prowadzonych zgodnie z metodą naukową. Jednakże, nawet wykorzystując EBM, pewne istotne aspekty wpływające na decyzję lekarza nadal można określić jako „ukryte”.

Rozwój medycyny opartej na faktach zbiegł się z czasem z gwałtownym postępowaniem w dziedzinie informatyki. Nie dziwi więc fakt, że dość szybko pojawiły się systemy mające w założeniu wspomagać lekarza w podejmowaniu decyzji. Określa się je mianem Systemów Wspomagania Decyzji Klinicznych (ang. *Clinical Decision Support Systems*). Przykładowe systemy tego typu to Gideon, Hepar czy, stworzony na Wojskowej Akademii Technicznej, System Wspomagania Diagnostyki (SWD) [2]. Pomimo zastosowania w nich nowoczesnych technologii, systemy tego typu nadal spotykają się z dużą rezerwą w środowisku lekarskim. Jednym z czynników jest tutaj fakt, że nie uwzględniają one (lub uwzględniają tylko w niewielkim stopniu) wspomnianej już wiedzy ukrytej. Drugim czynnikiem jest brak możliwości sterowania czułością i swoistością diagnozy [3]. Powoduje to niechęć lekarzy do tego typu systemów.

2 Uczenie głębokie

Przedstawione problemy systemów wspomagania decyzji medycznych mogłyby pomóc rozwiązać metody tzw. Uczenia Głębokiego (ang. *Deep Learning*). Algorytmy DL są silnie rozwijane w ostatnich latach, głównie za sprawą ich komercyjnych zastosowań przez firmy, takie jak Google, Microsoft czy Baidu. Metody te są inspirowane osiągnięciami w kognitywistyce, a zwłaszcza hipotezą „uczenia uniwersalnego”. Hipoteza ta głosi, że praktycznie wszystkie wysokopoziomowe procesy umysłowe są

nabywane za pomocą tego samego uniwersalnego algorytmu dla różnych obszarów mózgu.

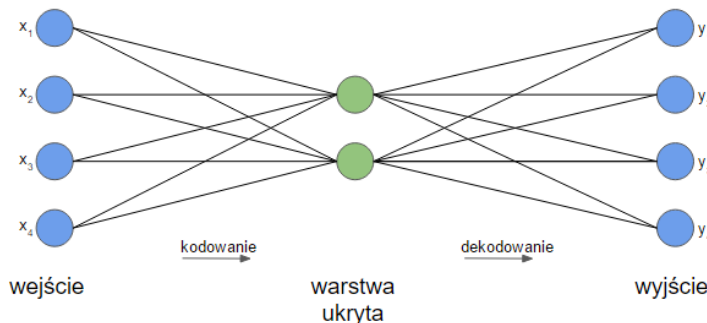
Mianem „uczenia głębokiego” określa się algorytmy uczenia maszynowego zdolne do samodzielnego wykrywania wysokopoziomowych abstrakcji w danych wejściowych. Abstrakcje te są często wykrywane przez algorytm samodzielnie, w sposób nienadzorowany. Cechą charakterystyczną jest wykrywanie cech w sposób iteracyjny, od oczywistych do tych bardziej subtelnych. Pozwala to algorytmom o stosunkowo prostej strukturze na nauczenie się nawet skomplikowanych pojęć.

Jedną z podstawowych architektur uczenia głębokiego są Stosowe Sieci Autoenkodujące (ang. *Stacked Autoencoders*) [4]. Autoenkoder to trójwarstwowa sieć neuronowa z liczbą neuronów wyjściowych równą liczbie neuronów wejściowych. Tym, co odróżnia autoenkoder od zwykłego modelu perceptronowego, jest algorytm uczenia, który w tym przypadku jest uczeniem nienadzorowanym: sieć uczy się odtwarzać sygnał podany na wejściu. Algorytm minimalizuje błąd rekonstrukcji sygnału, najczęściej w postaci błędu średniokwadratowego:

$$\mathcal{L}(x, x') = \|x - x'\|^2 = \|x - \sigma_2(W'\sigma_1(Wx + b)) + b'\|^2, \quad (1)$$

gdzie: x – wektor wejściowy; x' – wektor wyjściowy; σ_1, σ_2 – funkcje aktywacji; b, b' – wartości biasu; W, W' – macierze wag.

Jeżeli liczba neuronów w warstwie ukrytej jest mniejsza niż w warstwach zewnętrznych, wówczas sieć stara się wykryć takie cechy w danych, które umożliwiają efektywne kodowanie. Wykazano, że gdy neurony w warstwie ukrytej i wyjściowej mają liniowe funkcje aktywacji, to taka sieć uczy się aproksymować algorytm analizy głównych składowych. Schemat takiej sieci przedstawiono na rysunku 1.

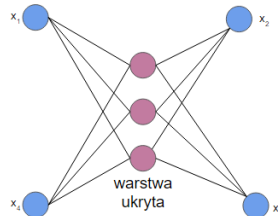


Rys. 1. Schemat sieci autoenkodującej; źródło: opracowanie własne

Fig. 1. Scheme of autencoding network; source: own construction

Do nauczonej sieci można następnie dołączyć kolejny autoenkoder w ten sposób, że warstwa ukryta poprzedniego staje się warstwą wejściową kolejnego. Stosując ten sam algorytm uczenia, kolejna warstwa uczy się ekstrakcji cech drugiego rzędu na podstawie cech zadanych na wejściu. Proces dołączania kolejnych warstw można powtórzyć wielokrotnie, uzyskując w wyniku stosowy autoenkoder. Ostatnim krokiem jest zazwyczaj dostrojenie ostatecznych wartości wag z wykorzystaniem standardowego algorytmu uczenia nadzorowanego, np. propagacji wstecznej.

Innym rodzajem architektury dla sieci głębokich są głębokie sieci przekonań (ang. *Deep Belief Networks*). Algorytm uczenia tych sieci, zaproponowany przez George'a Hinton, był jednym z pierwszych efektywnych algorytmów uczenia głębokiego [5].



Rys. 2. Schemat ograniczonej maszyny Boltzmanna; źródło: opracowanie własne
Fig. 2. Scheme of restricted Boltzmann machine; source: own construction

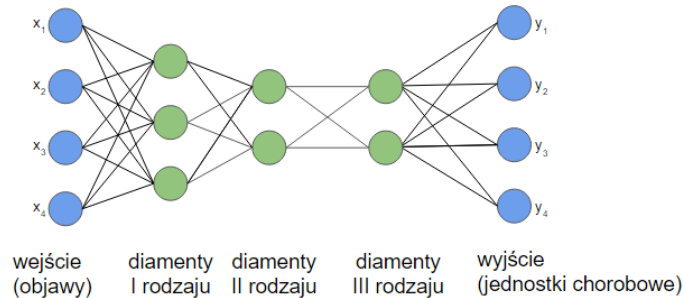
Podstawowym elementem składowym DBN jest ograniczona maszyna Boltzmanna (rys. 2). Maszyny Boltzmanna to sieci neuronowe, które są w stanie nauczyć się rozkładu prawdopodobieństwa danych wejściowych. W przypadku ograniczonych sieci pojawia się dodatkowy warunek: struktura sieci musi być grafem dwudzielnym, przy czym jedna grupa neuronów stanowi warstwę widoczną, a druga – ukrytą. Uczenie maszyny Boltzmanna należy do grupy metod bez nadzoru i polega na takiej modyfikacji wag W , aby zmaksymalizować iloczyn prawdopodobieństw dla pewnego stanu zbioru danych wejściowych X :

$$\arg \max_W \prod_{x \in X} P(x). \quad (2)$$

Algorytm tworzenia głębokiej sieci przekonań jest analogiczny jak w przypadku stosowych autoenkoderów. W tej sytuacji warstwa ukryta maszyny Boltzmanna staje się warstwą wejściową dla kolejnej maszyny, która również zostaje nauczona. Kolejne warstwy dokładane są iteracyjnie. Na koniec sieć jest dostrajana z użyciem uczenia nienadzorowanego lub nadzorowanego.

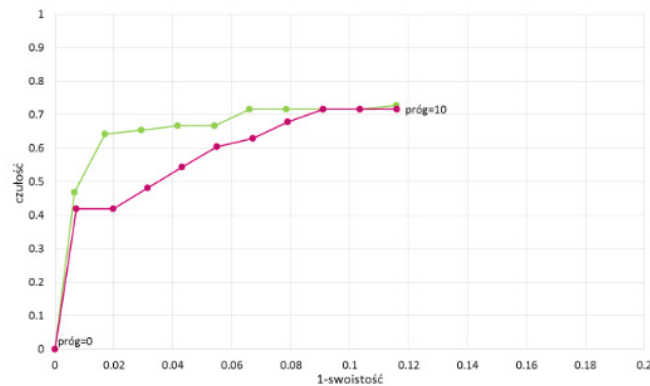
3 Uczenie głębokie a diagnostyka medyczna

Tym, co szczególnie predysponuje uczenie głębokie do zastosowania w medycynie, jest przede wszystkim zdolność do hierarchicznej ekstrakcji cech. Jak zauważyli Walczak i Paczkowski, odpowiednie przetworzenie danych medycznych pozwala wyodrębnić informacje na temat ważności symptomu lub zestawu symptomów, określane jako „diamenty diagnostyczne” [6]. Diament I rodzaju to objaw, który jednoznacznie wskazuje na jedną i tylko jedną jednostkę chorobową (określany jest on również mianem objawu patognomonicznego). Diament II rodzaju to objaw, który jest charakterystyczny dla pewnej jednostki chorobowej, ale występuje również w innych jednostkach chorobowych. Diament III rodzaju określa grupę objawów, która zawsze występuje w grupie objawów charakterystycznych dla jednostki chorobowej.



Rys. 3. Przykład głębokiej sieci dla diagnostyki medycznej; źródło: opracowanie własne
 Fig. 3. Example of deep network for medical diagnosis; source: own construction

Warto zauważyć, że diamenty diagnostyczne należą do cech o naturze hierarchicznej. Tradycyjne algorytmy klasyfikacji z reguły nie są w stanie wykorzystać tych danych, o ile nie są one podane w formie jawnej na wejściu klasyfikatora. Metody oparte na uczeniu głębokim byłyby zdolne do wykrycia tych cech samodzielnie, dla przykładu pierwsza warstwa ukryta wykrywałaby diamenty I rodzaju (objawy patognomoniczne), kolejna warstwa wykrywałaby na ich podstawie również diamenty II rodzaju, zaś ostatnia warstwa grupowałaby objawy w diamenty III rodzaju. Oczywiście, diamenty diagnostyczne to tylko przykład cech ukrytych, które mogłyby być wykrywane przez taki klasyfikator. Innymi cechami mogłyby być, dla przykładu, czynniki idiosynkratyczne lub zjawiska zaciemniające wyniki testów diagnostycznych.



Rys. 4. Zależność czułości i swoistości od wartości progu dla klasyfikatorów złożonych: metody Bordy (zielony) i najwyższej rangi (fioletowy); źródło: [3]
 Fig. 4. Relationship between sensitivity, specificity and rank threshold for classifier ensembles: Borda Count (green) and Highest Rank (purple); source: [3]

Innym zastosowaniem uczenia głębokiego w diagnostyce byłyby zdolność do sterowania parametrami klasyfikacji, takimi jak czułość, swoistość i próg odcięcia. Problem ten jest o tyle istotny, że parametry te są ze sobą silnie związane, co wykazano m. in. w [3].

Dotychczas nie powstała metoda analitycznego ustalania tych wartości dla dowolnego klasyfikatora. W przypadku uczenia głębokiego do istniejącego już klasyfikatora można dodać warstwy głębokie odpowiadające za modyfikację tych parametrów. Stanowiłoby zupełnie nową jakość w dziedzinie systemów wspomagania decyzji medycznych.

Literatura

1. Sackett D. L. i in.: Evidence based medicine: what it is and what it isn't, *BMJ*, nr 312, p. 71, 1996
2. Wojskowa Akademia Techniczna: *System Wspomagania Diagnostyki do informatycznego wspomagania badań naukowych w medycynie*, [Online] <http://www.isi.wat.edu.pl/sites/default/files/zalaczniki/swd.pdf>
3. Antczak K.: Rank Threshold in Classifier Ensembles in Medical Diagnosis, *Computer Science & Mathematical Modelling*, tom III, nr 1, 2016
4. Vincent P. i in.: Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion, *Journal of Machine Learning Research*, tom 11, pp. 3371-3408, 2010
5. Hinton G. E. i Salakhutdinov R. R.: Reducing the Dimensionality of Data with Neural Networks, *Science*, tom 313, nr 5786, pp. 504-507, 2006
6. Walczak A. i Paczkowski M.: Medical data preprocessing for increased selectivity of diagnosis, *BioAlgorithms and MedSystems*, tom 16, nr 1, March 2016

Streszczenie

W pracy przeanalizowano perspektywy zastosowania metod uczenia głębokiego w diagnostyce medycznej. Jedną z kluczowych cech uczenia głębokiego jest zdolność do wyodrębniania złożonych wzorców o strukturze hierarchicznej. Wzorce takie występują również w diagnostyce, jako tak zwane diamenty diagnostyczne. Zastosowanie głębokich sieci neuronowych mogłoby poprawić jakość klasyfikatorów wykrywających choroby na podstawie objawów. Dodatkowo umożliwiłoby to sterowanie czułością i swoistością klasyfikatorów.

Słowa kluczowe: sieci neuronowe, diagnostyka medyczna, uczenie głębokie

Deep Learning in Medical Diagnosis

Summary

In this paper we analyze perspectives of applying deep learning methods in a field of medical diagnosis. One of key features of deep learning is ability to extract complex, hierarchical patterns. Such patterns are present also in a medical diagnosis, where they are known as diagnostic diamonds. Applying deep neural networks could increase performance of medical classifiers. Moreover, it would allow to adjust sensitivity and specificity of classifiers.

Keywords: neural networks, medical diagnosis, deep learning