

MICHAŁ PODPORA

Wydział Elektrotechniki, Automatyki i Informatyki
Instytut Automatyki i Informatyki

HTM – NOWA ERA W HISTORII SZTUCZNYCH SIECI NEURONOWYCH?

Streszczenie: W tym artykule autor przybliży niedawno powstały, dość dobrze rokujący, kierunek badań nad dalszym rozwojem sieci neuronowych. Sieci HTM, dzięki temu, że całkowicie wzorowane są na strukturze i zasadzie działania kory nowej, otwierają nowe możliwości w tworzeniu maszyn „inteligentnych” w naszym ludzkim rozumieniu.

1. WSTĘP

Marzenia o tworzeniu istot /rzeczy inteligentnych towarzyszyły ludzkości od wieków. Wizerunek medialny „robotów” początku XX wieku odważnie przenikał się ze światem fantastyki. Istniała silna potrzeba rozpropagowania akceptacji, a nawet entuzjazmu w stosunku do rozwoju techniki w czasach, gdy w centrum uwagi była tematyka wynalezienia bomby atomowej [4]. Pokazywano coraz bardziej doniosłe „sukcesy” naukowców w postaci „robotów”, tak aby to właśnie one utożsamiały technikę. Gdy w 1947 roku w laboratoriach Bell’a zakończono pracę nad pierwszym tranzystorem, ludzkość ze zwiększoną siłą wierzyła w potęgę rozwoju własnej cywilizacji. Niestety rzeczywistość znacznie różniła się od tego, co podawały media – pokazywane roboty w żadnym przypadku nie przejawiały ani odrobiny inteligencji. Były to jedynie proste mechaniczne urządzenia, których sterowanie polegało na włączaniu i wyłączaniu (przez człowieka) poszczególnych silników lub podobnym rozwiązaniu.

Od ponad pół wieku naukowcy próbują stworzyć „inteligentną maszynę” – aż trudno uwierzyć, że tyle czasu minęło od opublikowanego w 1950 roku słynnego artykułu [1] Alana Turinga, w którym zaproponował, aby „inteligencję” maszyny oceniać nie na podstawie definicji, lecz za pomocą gry (testu). Takie podejście intuicyjnie wydaje się być najbardziej trafne – łatwiej jest zapytać człowieka czy rozmawia z człowiekiem, czy z maszyną, niż wymyślać rozmaite definicje i warunki, jakie musiałaby spełniać dana maszyna, by móc być uznaną za „inteligentną”. Jak się okazuje, człowiek bardzo dobrze sprawdza się w charakterze arbitra – potrafi natychmiast zauważyć najdrobniejszą anomalie czy choćby „dziwny styl wypowiedzi” rozmówcy (a są to zadania weryfikacji niezwykle trudne w ujęciu algorytmicznym).

W tym samym roku – 1950 – zostaje wydana legendarna powieść Isaaca Asimova „Ja, Robot”. Po raz pierwszy wydaje się, że technika rzeczywiście dogania świat fantastyki. Jedynie naukowcy zdają sobie sprawę z tego, że do prawdziwie „inteligentnych” maszyn jest jeszcze bardzo daleka droga.

Ponad pół wieku później stopień zaawansowania badań w zakresie sztucznej inteligencji nadal nie jest wystarczający. W miarę postępu (z wielkim trudem osiąganego) coraz bardziej zdumiewa potęgą ludzkiego umysłu.

Celem niniejszego artykułu przeglądowego jest przybliżenie czytelnikowi jednego z możliwych kierunków badań, obecnie aktywnie rozwijanego za oceanem i z powodzeniem stosowanego do zagadnień rozpoznawania i predykcji w wielu zastosowaniach (m.in. dotyczących ekonomii, biologii oraz wejść sensorycznych i interfejsów człowiek-maszyna).

2. SZTUCZNA INTELIGENCJA

Jedną z definicji terminu „sztuczna inteligencja” określa to pojęcie jako inteligencję maszyny uzyskaną w procesie inżynierskim (w przeciwieństwie do naturalnego). Jest to najbardziej pierwotna i mimo wszystko dość wąska definicja. Większość maszyn (lub aplikacji), o których potocznie mówi się, że „mają zaimplementowane metody sztucznej inteligencji”, faktycznie podlegają tej rdzennej definicji. Aby uwypuklić drobną różnicę między takim rozumieniem sztucznej inteligencji, a innymi przypadkami, można powiedzieć: „Nie jest ważne jak zaprojektowany jest sztuczny system, o ile tylko zachowuje się on tak jak człowiek” [3]. W tym podejściu istnieje analogia między obliczeniami a myśleniem. Mózg operuje abstrakcyjnymi symbolami bazując na stanie wejść tak samo jak każdy komputer – mózg jest więc rodzajem komputera [3]. W 1943 roku W. McCulloch i W. Pitts opublikowali pracę, w której opisali możliwość funkcjonowania pojedynczego neuronu jako bramki logicznej, co utwierdziło zwolenników porównywania mózgu z komputerem w przekonaniu, że badania są na właściwej ścieżce. W czasach powojennych, gdy komputery stawały się coraz bardziej dostępne, zaczęły pojawiać się nowe zastosowania i nowe implementacje. Cały świat nauki patrzył z nieskrywanym zainteresowaniem na rozkwit „sztucznej inteligencji” prorokując rychłe dorównanie możliwościom intelektualnym człowieka, a nawet ich przewyższenie. Tak się jednak nie stało. Zarówno tworzone maszyny, jak i pisane aplikacje, potrafiły rozwiązywać tylko takie zadania, jakie przewidział programista. Program był pisany pod kątem danego zadania – albo do tłumaczenia tekstu, albo do szachów, albo jakiegoś innego określonego celu. Program, który w maju 1997 roku wygrał z mistrzem świata Garrim Kasparovem w szachy, nie jest w stanie zagrać w warcaby. A i w szachy wygrał z mistrzem nie dla tego że był „mądrzejszy”, tylko dlatego, że był o wiele szybszy – człowiek „widzi”, w którym miejscu planszy rozegra się walka; komputer natomiast analizuje wszystkie możliwe kombinacje ruchów. Komputer potrafi wygrać w szachy z mistrzem „nie rozumiejąc” ich, podobnie jak kalkulator wykonuje działania „nie znając” matematyki [3]. Nawet jeden z najsłynniejszych programów w historii sztucznej inteligencji – Eliza – który w swoich czasach był niesamowicie blisko przejścia testu Turinga – działa w oparciu o bardzo prosty algorytm: wyszukuje w zdaniu słowa kluczowe i w oparciu o bazę zwrotów odpowiada – tylko wstawiając je w odpowiednim

miejscu frazy. W każdym przypadku jednak, algorytm danego programu postawiony w nowych, nieznanym warunkach, nie działa prawidłowo (lub częściej – nie działa wcale).

W pewnym momencie w historii sztucznej inteligencji pojawiło się nowe narzędzie – systemy eksperckie. W oparciu o predefiniowaną bazę danych z faktami, potrafią udzielać odpowiedzi na stawiane pytania. Systemy eksperckie okazały się bardzo przydatnym narzędziem i są stosowane z powodzeniem po dziś dzień, ale niestety nie przybliżają sztucznej inteligencji do ludzkiej.

3. SIECI NEURONOWE

Sieci neuronowe w latach 60-tych pojawiły się jako konkurencja dla zwolenników sztucznej inteligencji – konkurencja głównie w zakresie przyznawanych dotacji na badania naukowe. W latach 80-tych dobra passa sztucznej inteligencji zaczęła się mieć ku końcowi – niewykluczone, że w związku z brakiem jakichś spektakularnych osiągnięć, nowości. Zyskali na tym zwolennicy sieci neuronowych, a z biegiem czasu sieci neuronowe okazały się bardzo owocnym pomysłem.

Sieci neuronowe składają się z połączonych (w określoną strukturę) pojedynczych sztucznych neuronów, których sposób funkcjonowania jest całkowicie zaczerpnięty ze świata biologii. Eksperymenty ze sztucznymi sieciami neuronowymi pokazały, że dzięki odpowiednim połączeniom między neuronami, cała sieć jest w stanie rozwiązywać niektóre problemy nierozwiązywalne dla sztucznej inteligencji.

Budowa sieci neuronowej jest analogiczna do budowy mózgu – jedno i drugie tworzą odpowiednio połączone neurony, ponadto sieć neuronowa nie ma wydzielonego obszaru danych i listy instrukcji (jak jest w przypadku klasycznego programu komputerowego, a więc i algorytmu sztucznej inteligencji) Ciekawe jest, że w niektórych zastosowaniach sieć wystarczająca do otrzymania satysfakcjonującego wyniku składa się z kilku do kilkunastu neuronów – tak więc biorąc pod uwagę ilość neuronów w ludzkim mózgu (nikomu nie udało się jak na razie policzyć ile ich jest) różnica jest oczywista. Szacuje się, że kora nowa dorosłego człowieka ma około 30 miliardów neuronów. Wydaje się to być duża liczba, jeżeli jednak wziąć pod uwagę, że zapisane są tam wszystkie wspomnienia, doświadczenia, wiedza i uczucia, wówczas łatwiej jest zmienić zdanie na wręcz przeciwne – kora nowa zawiera „zaledwie” około 30 miliardów neuronów.

Skoro znana jest budowa neuronu, wiadomo jak łączyć neurony w sieć i jak później tą sieć nauczyć i odpytać, to czemu nie pokusić się o zbudowanie aplikacji „dorównującej” człowiekowi – czemu nie spróbować połączyć 30 miliardów neuronów? I czy to już będzie wtedy sztuczny mózg?

Zmysły, a dokładniej receptory, stanowią wejścia – dotyk, wzrok, słuch i pozostałe. Mózg jest natomiast zamkniętą w czaszce masą tkanki nerwowej, która bez tych wejść jest bezużyteczna. Mózg nie posiada wewnątrz nawet

receptorów (zupełnie nie poczuje ani ukłucia szpilką ani cięcia skalpelem). Podobnie jest ze sztuczną siecią neuronową – bez wejść nie znaczy nic. Natomiast czy sztuczna sieć neuronowa potrafi dorównać człowiekowi? Niestety nie – budowane sieci są albo zbyt małe, by „zobaczyć” wycinek otaczającego świata, albo zbyt duże, a co za tym idzie zbyt wolne i zbyt skomplikowane. Tym czasem biorąc pod uwagę, że typowy biologiczny neuron potrafi wygenerować potencjał czynnościowy i powrócić do poprzedniego stanu w czasie 5ms, można oszacować ile konkretnie neuronów bierze udział w danej operacji [5]. Załóżmy, że jakaś osoba ma do wykonania następujące zadanie: popatrzeć na zdjęcie, a jeśli na zdjęciu znajduje się pałac kultury, nacisnąć przycisk. Pamiętając o tym, jaki jest czas reakcji pojedynczego neuronu, można dojść do wniosku, że większość reakcji/czynności podejmowanych jest w zaledwie kilkudziesięciu – stu kilkudziesięciu krokach „algorytmu” (w literaturze [3] można spotkać tzw. „zasadę stu kroków”). Trudno powiedzieć ile czasu zajęłoby to komputerowi (i czy w ogóle byłoby to wykonalne bez specjalnego programu napisanego do tego konkretnego zadania). Niektórzy twierdzą, że ta miażdżąca przewaga mózgu nad komputerem wynika z tego, że mózg potrafi „przetwarzać” równoległe, podczas gdy komputery są maszynami sekwencyjnymi (dwa lub cztery rdzenie procesora też niewiele tu zmieniają wobec miliardów neuronów). Należy jednak pamiętać, że po wspomnianych stu krokach pojawia się wynik – reakcja odpowiednich mięśni. Trudno sobie wyobrazić algorytm (nawet całkowicie równoległy), który potrafiłby choćby zebrać informacje z 30 miliardów neuronów i odpowiednio przedstawić wyjścia w zaledwie stu krokach. Należy pamiętać, że w tym czasie (od bodźca do reakcji) zarówno komputer, jak i mózg musi wykonać mnóstwo operacji na danych wejściowych [6] : obraz przechwytywany (siatkówka / kamera) trafia (przez nerw wzrokowy / przez strumień wideo) do mózgu (lub komputera), gdzie rozmaite neurony (w ośrodku V1 / w funkcjach i procedurach przetwarzania obrazu) reagują na określone atrybuty obrazu i tak dalej. Skoro część czasu zostanie poświęcona przetworzeniu sygnału wejściowego, część przygotowaniu odpowiedzi na wyjściach, to ile kroków pozostaje do dyspozycji? Ile kroków trwa proces „zrozumienia” prezentowanej fotografii?

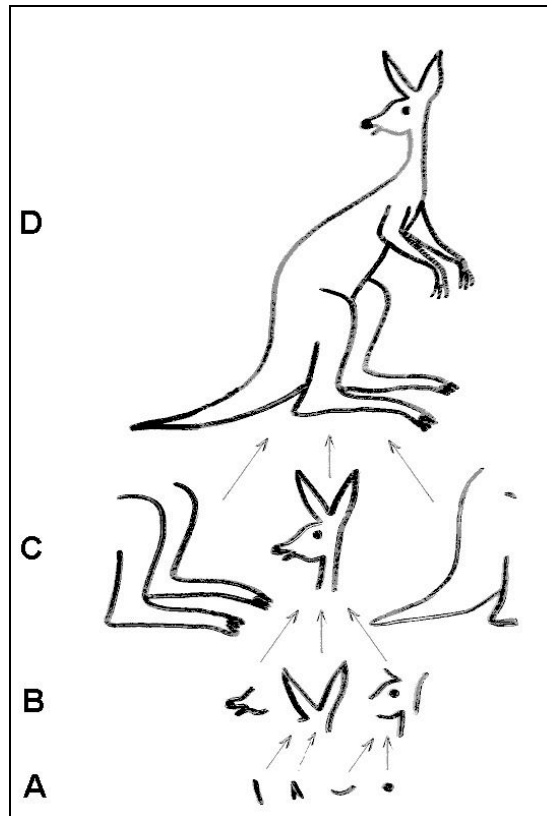
4. HTM

Jak to jest możliwe, że ewolucja potrafiła wytworzyć tak zaawansowane rozwiązanie jak mózg? Czy można zbudować maszynę, która byłaby uznana za „inteligentną” w naszym ludzkim, potocznym rozumieniu? Czy będzie można kiedyś nareszcie maszyny „uczyć” zamiast „programować”?

Od dawna już wiadomo, że kora nowa zbudowana jest z sześciu warstw neuronów. Wiadomo też, że ma grubość około 2 mm, że określone jej części zajmują się określonymi zadaniami. Naukowcy od lat potrafią powiedzieć, ile waży przeciętny mózg, jaką powierzchnię ma kora nowa, jak jest zbudowana, mimo to nadal nie potrafimy zbudować maszyny prawdziwie inteligentnej.

W 1978 roku Vernon Mountcastle opublikował artykuł, w którym stwierdził, że kora nowa ma „niezwykle jednolitą strukturę”. Obszary zajmujące się mową, słuchem, wzrokiem, dotykiem, sterowaniem mięśniami – wszystkie mają podobną strukturę. Oznacza to, że kora używa tego samego algorytmu do przetwarzania różnych danych wejściowych [2]. Jednak naukowcy zajęli się raczej wykorzystaniem sieci neuronowych, a nie poszukiwaniem tego „biologicznego” algorytmu. Tymczasem analizując zestawienie kilku neuronów nie można zrozumieć działania mózgu – tak samo jak eksperymentowanie kilkoma tranzystorami nie zaowocuje zrozumieniem funkcjonowania procesora. Sieci HTM noszą swoją nazwę nieprzypadkowo. Skrót ten pochodzi od angielskich słów „hierarchical temporal memory” i posiada następujące znaczenie [7]:

- Hierarchical – (hierarchiczne): Sieci HTM zbudowane są z węzłów połączonych hierarchicznie w topologii drzewiastej. Każdy z węzłów to algorytm, w którym zaimplementowano zarówno możliwość uczenia, jak i pamiętania danych wejściowych. Węzły niższej warstwy otrzymują dużą ilość danych wejściowych i po ich przetworzeniu wysyłają „opracowane” dane (na podstawie danych wejściowych) do warstwy wyższej (rysunek 1). Można więc powiedzieć, że każda warstwa sieci HTM ma zdolność tworzenia wyższego poziomu abstrakcji.
- Temporal – (czasowe): Podczas uczenia sieć HTM musi otrzymywać dane zmieniające się w czasie. O ile dla zmysłu słuchu lub dotyku zmiana bodźca (danych wejściowych) w czasie jest oczywista, to dla przetwarzania nieruchomego obrazu lub tabel z danymi statystycznymi należy tak przygotować dane wejściowe sieci HTM, aby były zmienne w czasie. Przykładowo nieruchomą klatkę obrazu wideo można na wejście sieci podawać w postaci przesuwanego się kolejnymi wierszami (o jeden piksel dalej) prostokąta o liczbie pikseli równej liczbie wejść sieci (w ten sposób sieć będzie „patrzeć” na kolejne fragmenty obrazu – wejścia sieci będą pobudzane danymi zmiennymi w czasie).
- Memory – (pamięć): Węzły sieci HTM działają w dwóch trybach, które można przyrównać do ćwiczenia pamięci (uczenie się) i odpytywania pamięci (tryb wnioskowania). Każdy z węzłów nauczonej sieci HTM „pamięta” określony fragment wzorca wejściowego (w przypadku węzłów warstwy najniższej) lub określony uogólniony zestaw wzorców współistniejących czasowo lub przestrzennie na wyższym poziomie abstrakcji (w przypadku węzłów warstw wyższych).



Rys. 1. Hierarchia wzorców na różnych poziomach abstrakcji [opracowanie własne]

5. HTM: HIERARCHICAL

Pomysł, w wyniku którego powstała idea sieci HTM, polegał na możliwie jak najbardziej wiernym odwzorowaniu „algorytmu” mózgu. Od lat wiadomo, że kora nowa składa się z sześciu warstw. Zadaniem niższych warstw przede wszystkim jest tworzenie nowych poziomów abstrakcji – *jeżeli patrzymy na coś, co ma ogon, wąsy, a obok leży mysz, to możliwe, że patrzymy na kota* – w przytoczonym przykładzie jest to po prostu suma cech obserwowalnych w przestrzeni (rzeczywistość może być bardziej skomplikowana). Najwyższa, czyli szósta warstwa kory nowej – tzw. warstwa kojarzeniowa – składa się ze stosunkowo niewielkiej ilości komórek, natomiast ogromnej ilości połączeń pomiędzy różnymi, nieraz dość odległymi, obszarami kory. Dzięki temu m.in. istnieje możliwość łączenia informacji pochodzących z różnych zmysłów – wówczas proces wnioskowania może być znacznie bardziej skomplikowany: Powiedzmy, że jeden z węzłów przedostatniej warstwy „stwierdza”, że „to co widać”, to na 80% jest pies, a na 20% kot. Inny węzeł tej warstwy „jest zdania”, że „to co słyhać”, to na 80% jest kwik świni, a na 20% miauknięcie kota [8].

Węzeł nadrzędny w warstwie kojarzeniowej porównuje te wyniki i może jasno określić, jaki jest ostateczny wynik wnioskowania.

Hierarchiczna organizacja węzłów w sieci HTM jest całkowicie wzorowana na najważniejszych połączeniach między warstwami występującymi w korze nowej. Najniższą warstwę tworzą obszary pierwszorzędowej kory sensorycznej, otrzymującej informacje z narządów zmysłów (dla wzroku obszar ten nazwano „V1”). Następnie informacje z V1 trafiają w górę hierarchii do innych obszarów (V2, V4, IT i do innych miejsc). Każdy z regionów przetwarza inny aspekt danych wejściowych – na innym poziomie abstrakcji. Przykładowo neurony w obszarze V4 reagują na kształty o średniej złożoności, na przykład kształty gwiazdek w kolorach niebieskim lub czerwonym [3]. Obszar nazwany MT reaguje na ruch obiektów. W wyższych warstwach kory wzrokowej (tu najważniejszy jest obszar IT) ulokowane są obszary związane z pamięcią złożonych obiektów, na przykład twarzy, zwierząt, przedmiotów. Po porównaniu informacji wizualnej z informacjami docierającymi z innych zmysłów (w warstwie kojarzeniowej) możliwe jest podjęcie reakcji. Pomimo tego, że zmysły są źródłami informacji i należy te informacje „tłumaczyć” na wyższe poziomy abstrakcji, a mięśnie są wyjściami informacji i abstrakcyjne pojęcia w korze należy „upraszczać” na pojęcia mniej abstrakcyjne (aż do „rozkazów” wysyłanych poszczególnym mięśniom), to algorytm działania kory pozostaje ten sam – jedynie kierunek przesyłania jest przeciwny (w dół hierarchii). Komunikacja tak naprawdę w obu przypadkach jest dwukierunkowa, różnica polega tylko na teoretycznym określeniu, które połączenia są połączeniami „nośnymi” informacji, a które są „zwrotnymi” (o połączeniach zwrotnych więcej w następnych rozdziałach).

Sieci HTM nie mają zdeterminowanej struktury. Do dyspozycji programisty-projektanta są węzły, które można łączyć ze sobą dowolnie, tworząc jakąkolwiek strukturę hierarchiczną. Węzły warstw wewnętrznych zazwyczaj projektuje się jako identyczne. Zdarzają się natomiast węzły, które mają odmienne wejścia lub wyjścia (np. wszystkie węzły pierwszej warstwy) – dla każdego rodzaju danych wejściowych powinien być wówczas stworzony odpowiedni „interfejs” wejściowy do sieci HTM – tzw. sensor. Przykładowo sensor sieci HTM przetwarzającej dźwięk mógłby zawierać algorytmy określające wysokość dźwięku i kształt fali akustycznej w strumieniu wejściowym. Następne warstwy używałyby tak skonwertowanej informacji do formowania kolejnych poziomów abstrakcji.

6. HTM: TEMPORAL

Istnieje kilka aspektów współistnienia czasu jako jednego z parametrów związanych z uczeniem czy też danymi źródłowymi. Najbardziej istotnym i czytelnym spośród nich, jest fakt reprezentacji danych wejściowych jako wzorców zmiennych w czasie. Najłatwiej to zrozumieć na podstawie przetwarzania dźwięku. W pierwszorzędowej korze A1 komórki nerwowe reagują na najbardziej „fizyczne” cechy słyszanego dźwięku. Tymczasem dane

wejściowe dochodzące do ośrodka A1 mogą być bardzo różne – każde słowo może być wypowiedziane z inną głośnością, szybkością, akcentem, barwą głosu. To właśnie dzięki podstawowej właściwości sieci HTM – zdolności do tworzenia wyższych poziomów abstrakcji – w wyniku określonej w czasie zmiany parametrów dźwięku możemy usłyszeć „słowo” (i skojarzyć jego znaczenie na przykład na podstawie wspomnień), „dźwięk” lub „melodię”. Ciekawym przykładem jest tutaj „melodia” – jest to określona w czasie kombinacja interwałów pomiędzy występującymi po sobie dźwiękami. Nie jest istotne, czy melodię ktoś nam zaśpiewa, czy zagra na flecie, czy na pianinie – przez ośrodek słuchu kory nowej zostanie rozpoznana ściśle określona melodia. Wbrew pozorom rozpoznawanie obrazu również może (a w sieciach HTM musi) mieć swój aspekt czasowy – gdy oglądamy zdjęcie, nasz wzrok nie jest utkwiony nieruchomo w środku fotografii, nie obejmuje również całego zdjęcia naraz. W rzeczywistości gałki oczne w ciągu każdej sekundy wykonują kilka drobnych (niezauważalnych dla świadomości) przemieszczeń zwanych sakkadami. Sakkady powodują, że na tzw. żółtą plamkę (miejsce o największej gęstości czopków na siatkówce oka) trafia coraz to inna część oglądanego zdjęcia. Można więc powiedzieć, że następny poziom abstrakcji określający oglądany przedmiot tworzony jest na podstawie zbioru cech z poprzedniego poziomu.

7. HTM: MEMORY

W korze nowej istnieje ogromna liczba połączeń zwrotnych między warstwami. Niektórzy naukowcy są zdania, że połączenia zwrotne stanowią więcej niż połowę wszystkich połączeń między neuronami. Po części spowodowane jest to mechanizmem uczenia/pamiętania – istnieje duże podobieństwo między „pamiętaniem” w komórkach kory nowej a typową pamięcią autoasocjacyjną. Po części natomiast wynika to z umiejętności „myślenia”, czyli świadomego pobudzania neuronów w oparciu o przechowywane „wspomnienia”. Jedną z zalet rozpoznawania wzorców przy użyciu pamięci autoasocjacyjnej, jest możliwość rozpoznania niekompletnych wzorców. Wynikającą z tego cechą „algorytmu” kory nowej, (związaną z istnieniem połączeń zwrotnych), jest zdolność przewidywania. Poszczególne neurony są pobudzane poprzez dane wejściowe wszystkich zmysłów – dotyczy to zarówno neuronów na poziomie sensorycznym, jak i na wyższych poziomach abstrakcji – „kompletując” widziany/słyszany/dotykany obiekt. W czasie pojawiania się kolejnych informacji o tym obiekcie, warstwa kojarzeniowa już może mieć część informacji wystarczającą do określenia, jaki to obiekt. Wystarczy usłyszeć psa, by „mieć przed oczami” jego wielkość, by określić czy jest groźny czy nie. To właśnie dzięki najwyższej warstwie – kojarzeniowej – możemy mówić o istnieniu tzw. niezmiennych reprezentacji [3], czyli abstrakcyjnych pojęć, określanych przez stopień pobudzenia neuronów z niższych warstw.

8. NIEZMIENNE REPREZENTACJE

Neurony w najniższych warstwach – wszystko jedno, czy przetwarzają obraz, dźwięk czy dotyk – otrzymują bardzo proste informacje: natężenie (jasności/ koloru/ dźwięku/ nacisku) w określonym punkcie (wzrok: przestrzeni; dźwięk: częstotliwości; dotyk: w określonym miejscu ciała). Wyższe warstwy grupują wyjścia tych neuronów (patrz: rysunek 1) i reagują na bardziej złożone (abstrakcyjne) wzorce przestrzenne i/lub czasowe (wzrok: linie, zgięcia; dźwięk: sylaby, nuty). W kolejnych warstwach tworzone są jeszcze bardziej abstrakcyjne pojęcia (wzrok: np. gwiazdki; słuch: np. melodia „sto lat”). Podążając w górę hierarchii można na przykład (w obszarze zwanym „IT”) znaleźć neurony reagujące na pojawienie się twarzy w polu widzenia. W przeciwieństwie do neuronów z najniższych warstw, które często zmieniają swój stan, neurony w wyższych warstwach wysyłają serie sygnałów tak długo, jak istnieje dany bodziec – np. „twarz” [3]. Pozwala to neuronom najwyższych warstw na wspomniane „skompletowanie” określonego abstrakcyjnego pojęcia (czyli niezmiennych reprezentacji). Wynik jest taki, że jeżeli widzimy psa, ale on nie szczeka, to niezmienna reprezentacja przechowywana w korze nowej podpowiada nam, że gdyby wydał z siebie jakiś dźwięk, to pewnie będzie to szczekanie. Mózg potrafi perfekcyjnie operować niezmiennymi reprezentacjami. Dla przykładu rozumienie mowy prawdopodobnie nie byłoby możliwe (albo byłoby bardzo utrudnione) gdyby nie ta umiejętność kory. Większość słów, które „docierają” do naszych uszu jest na tyle zniekształconych/ /zakłóconych/niekompletnych, że zrozumienie ich znaczenia możliwe jest jedynie przy użyciu niezmiennych reprezentacji. Czasem niezmiennie reprezentacje mogą się przydać również na wyższym poziomie abstrakcji – gdy chcemy zrozumieć sens zdania, podczas gdy nie słyszeliśmy kilku wyrazów.

Pole zastosowań niezmiennych reprezentacji jest ogromne. Najprostszym przykładem są programy rozpoznawania tekstu (tzw. OCR). Niektóre z nich korzystają z sieci neuronowych i/lub baz wiedzy w postaci słowników. Niestety żaden program nie jest lepszy od człowieka, ponieważ w sytuacji, gdy określona litera jest nieczytelna, powstaje dylemat, jakie słowo powinno być ostatecznym wynikiem przetwarzania. Metodami słownikowymi można wykluczyć kombinacje liter nie tworzące wyrazów, ale gdy mimo to zostanie kilka możliwości, to którą wybrać? Najgroźniejszym przykładem jest sytuacja, gdy dana litera zostanie omyłkowo rozpoznana jako inna i – zupełnie przypadkowo – okaże się, że istnieje taki wyraz. Nawet procedura zaznaczania „niepewnych fragmentów” może tego nie zauważyć. Dla człowieka (lub programu korzystającego z sieci HTM) sytuacja taka nie miałaby miejsca właśnie dzięki niezmiennym reprezentacjom i mechanizmom przewidywania i uzupełniania. Gdyby w zwrocie „części mowy” literka „o” została błędnie rozpoznana jako „e” („części mewy”) słownikowe sprawdzenie nie zauważyłoby żadnej nieścisłości, podczas gdy niezmienna reprezentacja związana ze słowem „mewa” określa zwierzę (które nijak nie pasuje do pierwotnego kontekstu zdania). Dlatego można śmiało stwierdzić, że to właśnie

w tym podejściu – w sieciach HTM – leży klucz do tworzenia prawdziwie inteligentnych programów i maszyn.

9. DLA KOGO SIECI HTM?

Pomimo tego, że sieci HTM są stosunkowo świeżym pomysłem, dość szybko zaczęły zdobywać swoich zwolenników. Praktycznie w każdej branży, w której przydatne są aplikacje oparte o sieci neuronowe, sieci HTM mogą się okazać bardzo pożyteczne. Szczególnie obiecujące są zastosowania, w których pożądana jest analogia pomiędzy działaniem aplikacji i ludzkiego mózgu, ale to nie jest jedyne pole zastosowań. Najbardziej zaangażowana instytucja – firma Numenta – oferuje oprócz specjalnie przygotowanych narzędzi (dostępnych pod specjalną licencją, nieodpłatnie dla celów niekomercyjnych, w tym naukowo-badawczych) zarówno dość obszerną dokumentację swojego produktu (platformy programistycznej sieci HTM), jak i artykuły na temat sieci HTM i problematyki, która „pasuje” do tego rozwiązania [9]. Wśród przykładów dołączonych do dystrybucji platformy można znaleźć nie tylko zastosowania związane z wzrokiem czy słuchem, ale np. program analizujący ciągi znaków w pliku, czy program analizujący przepływ wody w rzece. Nie zmienia to faktu, że najbardziej spektakularnym programem w roku 2007 był program „Pictures” [10] rozpoznający narysowany przez użytkownika kształt spośród 48 wcześniej nauczonych kształtów. Ta skromna aplikacja działająca pod systemem Microsoft Windows w prosty i dobitny sposób prezentuje przewagę sieci HTM nad sieciami neuronowymi – jeżeli użytkownik narysuje „kubek” z uchem po lewej stronie (podczas gdy sieć HTM uczona była wyłącznie rysunkami kubków z uchem po prawej stronie), to ten rysunek i tak zostanie rozpoznany jako „kubek”. Jeżeli wspomniany kubek zawsze był wielkości całego okna aplikacji, a użytkownik tym razem narysował mały kubeczek w rogu, też zostanie prawidłowo rozpoznany. Dla sieci neuronowych te dwa proste przypadki już stanowiłyby pewnego rodzaju wyzwanie (programista musiałby przewidzieć wszystkie takie sytuacje przy projektowaniu i uczeniu sieci), natomiast w przypadku sieci HTM prawidłowy wynik rozpoznawania pojawia się naturalnie, dzięki jej zasadzie funkcjonowania – cechy obserwowanego obrazu łączą się w logiczne grupy, które na odpowiednich poziomach abstrakcji mają swoje reprezentacje sumujące się w niezmienną reprezentację danego obiektu (rysunek 1). Kot ma głowę, ogon, łapki, wąsiki – nie jest istotne, czy głowa jest z lewej strony zdjęcia, a ogon z prawej, czy odwrotnie – nadal jest to kot (tyle, że ze zmianą występowania wzorców w czasie).

10. WNIOSKI

Pięcioletnie dziecko rozumie i samodzielnie posługuje się językiem mówionym, potrafi odróżnić psa od kota, potrafi grać piłką – te trzy umiejętności (tak jak i wiele innych) są proste dla ludzi, a jednocześnie niewykonalne dla robotów [2]. Po części sytuacja ta wynika z tego, że naukowcy przez pół wieku starali się tworzyć inteligentne maszyny w oparciu

o logikę komputera. Profesor filozofii John Searle aby udowodnić, że komputery nie mogą być inteligentne, zaproponował w 1980 roku eksperyment myślowy zwany „chińskim pokojem” [3] – osoba mówiąca wyłącznie po angielsku, zamknięta w pokoju z biurkiem, otrzymuje opowiadanie i dotyczące go pytania, wszystko napisane po chińsku. Człowiek ten ma do dyspozycji stosy kartek papieru, ołówki i księgę z instrukcjami. Czasem instrukcje każą napisać pewne znaki na kartkach, czasem zamienić ich kolejność lub wymazać niektóre z nich. Kiedy instrukcje stwierdzą, że to już koniec, na kartce znajduje się jakiś zestaw znaków, który jest odpowiedzią na pytania. Osoba, która zna chiński widzi, że wszystkie odpowiedzi są poprawne, a niektóre nawet błyskotliwe. Tylko czy poprawność i jakość tych odpowiedzi wynika z inteligencji osoby, która odpowiadała na pytania? „Chiński pokój” wywołał prawdziwą burzę między zwolennikami i przeciwnikami sztucznej inteligencji – trudno zdefiniować kryteria, kiedy dany system jest inteligentny, a kiedy nie. Jeżeli za miarę inteligencji maszyny określimy stopień podobieństwa jej przetwarzania do procesów myślowych człowieka, wówczas możemy śmiało powiedzieć, że inteligentne maszyny jeszcze nie istnieją. Świat opisywany w powieściach Isaaca Asimova był, jest i jeszcze bardzo długo będzie fantastyką. Nikt nie wie, czy sieci HTM okażą się ważnym krokiem w stronę stworzenia inteligentnych maszyn – ale jest to następny krok. Roboty już dziś stanowią bardzo cenioną siłę roboczą [4], może kiedyś staną się czymś więcej niż narzędziami czy zabawkami. Historia rozwoju robotyki dopiero się zaczyna.

LITERATURA

- [1] TURING A.M., Computing machinery and intelligence, Mind, nr 59, strony 433-460, 1950
- [2] HAWKINS J., Learn Like a Human, IEEE Spectrum, nr04/2007
- [3] HAWKINS J., BLAKESLEE S., On Intelligence, Times Books, 2004
- [4] National Geographic – Frontlines of Construction – Construction Robots, National Geographic Channel, 2000
- [5] KURZWEIL R., The Age of Spiritual Machines: When Computer Exceed Human Intelligence, Penguin Books, 2000
- [6] PODPORA M., Computer vision in parallel computing, Przegląd Elektrotechniczny, strony 68-70, nr11/2007
- [7] Numenta, What's in a Name: Hierarchical Temporal Memory, Internet (dostęp: 2007-05-01): <http://www.numenta.com/about-numenta/numenta-technology.php>, 2007
- [8] HAWKINS J., DILEEP G., Hierarchical Temporal Memory – Concepts, Theory and Terminology, Internet (dostęp: 2007-05-01): http://www.numenta.com/Numenta_HTM_Concepts.pdf, 2007
- [9] Numenta, Problems that Fit HTMs whitepaper, Internet (dostęp: 2007-05-01): <http://www.numenta.com/for-developers/education/ProblemsThatFitHTMs.pdf>, 2007

- [10] Numenta, Aplikacja Pictures, Internet (dostęp: 2007-05-01):
<http://www.numenta.com/about-numenta/technology/pictures-demo.php>

Abstract:

In this article author presents a new, promising direction of further research into artificial neural networks and artificial intelligence. HTM networks are based on the structure and principles of functioning of neocortex, therefore they offer new possibilities creating intelligent machines.