

ANDRZEJ NOWAK, MICHAŁ ŻOCHOWSKI

*Instytut Studiów Społecznych Uniwersytetu Warszawskiego
Stawki 5/7, 00-183 Warszawa*

AFEKT W MODELACH SIECI NEURONOWYCH

W artykule tym przedstawiamy ogólną koncepcję zagadnienia, jak w sztucznych sieciach neuronowych można modelować współdziałanie procesów emocjonalnych (zwanymi procesami afektywnymi) z procesami poznawczymi. Przedstawione modele są oparte na empirycznej wiedzy z zakresu psychologii, dotyczącej relacji między emocjami a poznaniem, oraz na wykorzystaniu faktu, że z punktu widzenia fizyki sieci neuronowe są układami dynamicznymi. W modelach procesy emocjonalne kontrolują

przebieg procesów poznawczych, zmieniając w trakcie rozpoznania właściwości dynamiki sieci neuronowej. Modele z tak wbudowanymi mechanizmami kontroli funkcjonują sprawniej, niż równoważne modele nie uwzględniające wpływu emocji na poznanie. Ważnym wynikiem symulacji komputerowych, opisanych poniżej modeli, jest pokazanie, jak procesy emocjonalne mogą zwiększać możliwości przetwarzania informacji przez mózg.

EMOCJE A POZNANIE

W definicji gatunku *Homo sapiens* zdolność do myślenia abstrakcyjnego stanowi najważniejszą, wyróżniającą wręcz, właściwość człowieka. Badanie procesów myślenia a szerzej procesów poznawczych stanowiło przez ostatnie prawie trzydzieści lat jeden z najważniejszych obszarów zainteresowania psychologii. Stworzono precyzyjne modele najróżniejszych procesów poznawczych, takich jak spostrzeganie, rozpoznawanie wzorów, uczenie się, zapamiętywanie, rozumowanie czy podejmowanie decyzji. Wypracowano procedury eksperymentalne pozwalające doświadczalnie sprawdzać adekwatność modeli. Psychologia poznawcza jest wręcz uważana za najbardziej rozwiniętą dziedzinę psychologii.

Procesy emocjonalne właściwie nie były uwzględniane przy tworzeniu modeli procesów poznawczych. Koncentracja badań na procesach poznawczych wynikała, po pierwsze, z przeświadczenia o tym, że są one dużo ważniejsze od procesów emocjonalnych dla orientacji w świecie i że to one głównie decydują o zachowa-

niu. Po drugie, z przekonania o pierwotności procesów poznawczych wobec emocji (LAZARUS 1982). Zgodnie z tą koncepcją emocje są konsekwencją procesów poznawczych, są więc wobec procesów poznawczych wtórne. Dochodzące do mózgu informacje najpierw są przetwarzane przez system poznawczy: są rozpoznawane obiekty, nadawane jest znaczenie zdarzeniom, są wydobywane z pamięci informacje. To jakie emocje zostaną wzbudzone zależy od przebiegu procesów poznawczych. Dopiero wtedy gdy rozpoznamy, że zwierzę zagrażające nam drogę to tygrys, lub gdy zrozumiemy, że widziana przez nas rura to lufa wymierzonego w nas pistoletu zaczynamy się bać. Dopiero wtedy gdy rozpoznamy, że widziany przez nas owoc to cytryna zaczniemy się ślinić. Ponieważ emocje, zgodnie z tą koncepcją, pojawiają się dopiero po tym, jak zostaną przetworzone informacje przez system poznawczy, można je pominąć przy wyjaśnianiu przebiegu procesów poznawczych, gdyż to poznanie decyduje o ich kształcie.

ZNACZENIE PROCESÓW EMOCJONALNYCH

W ostatnich latach badacze przypisują coraz większą rolę emocjom. W licznych ekspery-

mentach psychologicznych wykazano, że emocje mogą powstawać zanim system poznawczy

przetworzy informacje i że raz wzbudzone mogą być niezależne od wyników przetwarzania przez system poznawczy (ZAJONC 1980, ZAJONC 1997). Wyrażane jest wręcz przekonanie, że procesy emocjonalne są dla funkcjonowania człowieka ważniejsze od procesów poznawczych (DAMASIO 1994, EDELMANN 1992, ZAJONC 1997). Zgodnie z tą koncepcją emocje mogą funkcjonować bez udziału procesów poznawczych, natomiast te ostatnie silnie zależą od procesów emocjonalnych i są w znacznej mierze w stosunku do nich wtórne.

ZAJONC (1997) twierdzi, że emocje są najważniejszą klasą procesów psychologicznych. Zdolność rozróżnienia między zdarzeniami pozytywnymi i negatywnymi jest według niego najbardziej podstawowym wymogiem wobec organizmu. Bez procesów prowadzących do wartościowania nie mogłoby bowiem istnieć życie. Tendencje do selektywnego zbliżania się lub unikania są obecne u wszystkich zwierząt. Procesy uczenia się i warunkowania zachodzą tak, by zwiększyć szansę występowania zdarzeń pozytywnych czyli nagród, a zmniejszyć szansę

występowania zdarzeń negatywnych czyli kar. Kategorie poznawcze nie grupują wyłącznie zdarzeń i obiektów podobnych do siebie, lecz w znaczącym stopniu grupują te zdarzenia i obiekty, które mają podobne znaczenie dla organizmu.

Na gruncie neurofizjologii EDELMANN (1992) wyróżnia dwa systemy w mózgu. System wartości kontroluje takie funkcje jak: apetyt, zachowania seksualne i zachowania agresywne. Funkcją tego systemu jest dbanie o to, by były zaspokojone potrzeby organizmu. System ten jest w swej definicji zbliżony do rozważanej przez nas funkcji emocji. System poznawczy służy głównie kategoryzacji bodźców. EDELMANN twierdzi, że główną funkcją systemu poznawczego jest wspomaganie systemu wartości i bez jego udziału system poznawczy nie mógłby się rozwinąć. Wykazuje on zarówno na gruncie wyników doświadczeń neurofizjologicznych, jak i symulacji komputerowych, że system poznawczy samoorganizuje się pod wpływem systemu wartości tak, by organizm mógł zdobywać rzeczy korzystne a unikać szkodliwych.

NATURA EMOCJI

Wyniki eksperymentów i koncepcje teoretyczne w psychologii wskazują, że system emocjonalny i system poznawczy stanowią dwa odrębne, lecz wpływające na siebie nawzajem systemy. Procesy poznawcze i emocje różnią się od siebie w sposób zasadniczy. Istotą procesów poznawczych jest przetwarzanie informacji. System poznawczy kategoryzuje nadchodzące bodźce, rozpoznaje obiekty, zapamiętuje i wydobywa z pamięci, wnioskuje i przewiduje. ZAJONC (1997), unikając wprowadzenia bezpośredniej definicji emocji, definiuje zdolność do reakcji emocjonalnych jako zdolność do reagowania adaptacyjnego i wartościującego na obecne i przewidywane warunki, które prawdopodobnie są szkodliwe, lub korzystne dla jednostki lub jej otoczenia społecznego. Ważną charakterystyką reakcji emocjonalnych jest zatem ich aspekt wartościujący.

Przyjmuje się w psychologii (PLUTCHIK 1962, 1980, por. REYKOWSKI 1972), że emocje można scharakteryzować w trzech wymiarach. Po pierwsze można określać ich intensywność, czyli odpowiadający im stan pobudzenia. Na jednym

krańcu są stany relaksu, odprężenia, senności, na drugim stany wysokiego pobudzenia, przeżycia, wściekłości, ekstazy. Stopień pobudzenia odnosi się do mobilizacji rezerw organizmu. Mówi się też o energetycznym aspekcie emocji. Drugi wymiar odnosi się właśnie do wartościowania. Jego krańce określają „dobry” i „zły”, bądź „przyjemny” i „nieprzyjemny”. Negatywne reakcje wartościujące, jak na przykład strach czy wstręt, wywołują tendencję do unikania a pozytywne, na przykład radość, wywołują tendencję do zbliżania się do wywołującego je bodźca. Wymiar ten jest szczególnie ważny zarówno dla ustosunkowania się wobec środowiska, jak też dla procesów uczenia się. Bodźce, które wywołują stany pozytywne działają jako nagrody i powodują utrwalanie się reakcji, te które wywołują stany negatywne działają jako kary. Trzeci wymiar to jakość czy treść emocji. Chociaż strach, smutek, wstręt lub wściekłość są emocjami o negatywnym znaku, to one są różne od siebie treścią i skłaniają do różnych zachowań.

ZAGADKA WCZESNEGO ROZPOZNANIA

Jednym z najbardziej intrygujących faktów dotyczących emocji jest to, że reakcja emocjonalna może być reakcją pierwotną, emocje mogą powstawać w odpowiedzi na bodziec, zanim

zostanie on rozpoznany przez system poznawczy. Istotę tej zagadki łatwiej zrozumieć zestawiając wyniki doświadczeń ZAJONCA (1980, 1997) ze stanowiskiem LAZARUSA (1982), który

reprezentuje klasyczny punkt widzenia. LAZARUS twierdzi, że poznanie jest procesem pierwotnym i wystąpienie jakichkolwiek reakcji emocjonalnych na wzorzec musi być poprzedzone rozpoznaniem tego wzorca przez system poznawczy. Jak inaczej bowiem system afektywny mógłby wiedzieć, na co ma reagować? Skąd człowiek może wiedzieć czy bać się czy cieszyć zanim nie rozpozna tego, co widzi? Z logicznego punktu widzenia rozpoznanie musi wyprzedzić reakcję emocjonalną.

Wyniki licznych eksperymentów potwierdzają jednak stanowisko Zajonca (ZAJONC 1980, 1997, KUNST-WILSON i ZAJONC 1980, MURPHY i ZAJONC 1993). Wskazują one, że reakcja emocjonalna może powstać niezależnie od reakcji poznawczej, a nawet może wyprzedzić ją w czasie. W eksperymentach badano różnice między reakcją emocjonalną na wzorce znane i wzorce nowe. Wykazano, że wzorce znane wywołują pozytywną reakcję emocjonalną, bodźce nieznane zaś wywołują reakcję negatywną. Innymi słowy lubimy to, co znamy, nie lubimy zaś rzeczy nieznanych. W typowym eksperymencie (KUNST-WILSON i ZAJONC 1980) najpierw pokazywano osobom badanym pewną liczbę chińskich znaków, zakładając, że osoby te nie znają tych znaków. W drugiej części osobom tym pokazywano parami jeden ze znaków wcześniej widzianych oraz inny, nieznanymi chiński znak. Badani mieli oceniać, który ze znaków bardziej im się podoba oraz rozpoznać, który ze znaków wcześniej widzieli. Znaki pokazywane w pierwszej części eksperymentu były znacznie pozytywniej oceniane, nawet jeżeli osoby nie potrafiły ich rozpoznać jako wcześniej widzianych. Wzorce wcześniej widziane wywoływały bardziej pozytywną reakcję nawet wtedy, jeżeli w pierwszej części eksperymentu pokazywano je na ekranie tak krótko, że nie można było rozpoznać ich kształtu. Reakcja emocjonalna, jak wskazują również wyniki wielu innych eksperymentów, nie jest zależna od rozpoznania bodźca przez system poznawczy, a nawet wydaje się wyprzedzać w czasie rozpoznanie, innymi słowy układ emocjonalny włącza się zanim układ poznawczy dokonał całkowitego rozpoznania bodźca. Reakcja emocjonalna występuje bardzo szybko, jest ona wrażliwsza od rozpoznania, tak że reagujemy bardziej pozytywnie na wzorce tak słabo pamiętane, że system poznawczy nie jest w stanie ich rozpoznać.

Selektywna wczesna reakcja na bodźce nowe leży też u podłoża odruchu orientacyjnego (SOKOLOV 1963, 1969), którego głównym komponentem jest skierowanie uwagi na nowy bodziec. Pojawienie się nowego bodźca powoduje skierowanie nań uwagi, wzrok skierowany zo-

staje w tę stronę, zwierzęta stawiają uszy, wzrasta poziom aktywacji centralnego układu nerwowego. Odruch orientacyjny pomaga skierować rezerwy systemu poznawczego tam, gdzie jest to najbardziej potrzebne czyli do przetworzenie nowej informacji. Ciekawe jest to, że odruch orientacyjny jest również wywoływany przez brak w bodźcach elementów normalnie w nich występujących. Jest więc on reakcją na nowość konfiguracji cech, a nie tylko na pojawienie się elementów uprzednio nieznanymi.

Dane doświadczalne wskazują również na to, że jest możliwe określenie treści emocjonalnej wzorca przed jego rozpoznaniem. POSNER (1981) wykazał, że gdy bardzo krótko pokazujemy na ekranie słowa zabarwione emocjonalnie, to osoba badana, nie będąc w stanie odczytać słowa, częściej niż by to wypadało losowo mylnie podaje inne słowo o zbliżonej treści emocjonalnej. MURPHY i ZAJONC (1993) wykazali, że bardzo krótkie, tak zwane podprogowe, wyświetlenie na ekranie fotografii twarzy uśmiechniętej lub wyrażającej ból wywołuje wyraźną zmianę reakcji emocjonalnej, mimo że osoby badane w ogóle nie zdają sobie sprawy z tego, że widzą jakąkolwiek twarz.

We wszystkich przypadkach omówionych powyżej zastanawia zjawisko, które można określić jako przedrozpoznanie bodźca przez system emocjonalny. Zanim zostanie on rozpoznany przez system poznawczy, w jakiś sposób może być wstępnie klasyfikowany, na co wskazuje występowanie selektywnej reakcji emocjonalnej. Takie przedrozpoznanie w postaci reakcji emocjonalnej jest łatwe do wytłumaczenia z funkcjonalnego punktu widzenia. Jeżeli coś dobrze znam i żyję, to oznacza to, że to coś nie jest zwykle śmiertelnie groźne; rzeczy nieznanne mogą zaś być bardzo niebezpieczne. Jeżeli w moim polu widzenia pojawi się nowy bodziec, to nim przede wszystkim powinienem się zająć, gdyż jego pojawienie się może wymagać szybkiej reakcji, może on bowiem stanowić zarówno zagrożenie, jak i potencjalne pożywienie.

Łatwo jest wytłumaczyć, dlaczego człowiek zobaczywszy drapieżnika najpierw powinien zacząć uciekać, a dopiero potem zastanowić się, czy to był lew czy tygrys. Nie jest natomiast jasne wytłumaczenie, jak jest możliwa selektywna reakcja emocjonalna na bodziec bez rozpoznania tego bodźca. Innymi słowy, skąd człowiek może wiedzieć, że bodziec jest nowy, zanim nie spróbował go rozpoznać? Skąd człowiek może wiedzieć, że powinien się bać, zanim rozpozna, że jest to tygrys? Na pytania te bardzo trudno odpowiedzieć w tradycyjnych modelach umysłu. Wygląda to wręcz na paradoks, tak jakby skutek rozpoznania miał poprzedzić w

czasie przyczynę, czyli proces rozpoznania. Prace prowadzone w naszym zespole wykazały, że na pytanie to, które w tradycyjnych modelach umysłu wydaje się mieć charakter paradoksu, można stosunkowo łatwo odpowiedzieć w modelach sztucznych sieci neuronowych, nowej klasie modeli poznania.

Przekonanie o pierwotnej naturze reakcji emocjonalnych zgodne jest też ze współczesnym poglądem na charakter związku emocji z przetwarzaniem informacji. Związek ten jest dwukierunkowy. To że procesy poznawcze mogą wywoływać emocje jest oczywiste. Jeżeli na przykład przeczytam w atlasie, że grzyb, który przed chwilą zjadłem, może być trujący, doświadczę lęku. Jeżeli znajdę wyjście z trudnej sytuacji, doświadczę uczucia ulgi lub radości. Ciekawszy wydaje się drugi kierunek tego związku, wpływ emocji na procesy poznawcze. W licznych badaniach wykazano, że emocje silnie wpływają na przetwarzanie informacji, można je wręcz traktować jako mechanizmy kontroli przebiegu procesów poznawczych (CARVER i SCHEIER 1990, OATLEY i JOHNSON-LAIRD 1987). Negatywne emocje mogą być traktowane jako komenda przerwania wykonywanego dzia-

łania lub zachodzącego procesu poznawczego (SIMON 1967). Pojawienie się negatywnej emocji oznacza zgodnie z tą koncepcją, że jest wykonywana niewłaściwa czynność poznawcza, którą należy przerwać i zaplanować nową.

Innym efektem wpływu emocji na poznanie jest zgodność między treścią emocji a treścią poznania. Gdy jesteśmy smutni, to smutne myśli przychodzą nam do głowy, przypominają się niepowodzenia, porażki i pogrzeby. Gdy jesteśmy zadowoleni, to przypominają się sukcesy, świat jest pełen możliwości a ludzie przyjaźni. Przy założeniu, że emocje są czynnikami kontroli procesów poznawczych można to wyjaśnić przyjmując, że emocje ustawiają system poznawczy w jednym z kilku jakościowo od siebie różnych trybów, włączając pewne mechanizmy poznawcze a wyłączając inne (OATLEY i JOHNSON-LAIRD 1987). Tak więc system poznawczy działa inaczej, gdy na przykład odczuwa się smutek, niż wtedy gdy odczuwa się złość. Gdy dominuje złość są tworzone plany ataku lub obrony, inne osoby są dzielone na sprzymierzeńców i wrogów, przedmioty spostrzegane jako potencjalna broń.

ATRAKTOROWA SIEĆ NEURONOWA — PROGRAMOWALNY UKŁAD DYNAMICZNY

Modele sztucznych sieci neuronowych w ostatnim dziesięcioleciu przeżywają rozkwit, zarówno jako dziedzina badań naukowych, jak też jako narzędzie znajdujące wiele zastosowań praktycznych (por. RUMELHART i MCCLELLAND 1986, HERTZ i współaut. 1991). Badania sztucznych sieci neuronowych są prawdziwie interdyscyplinarną dziedziną wiedzy. Dla psychologów poznawczych i neurofizjologów są one fascynujące, gdyż pokazują, jak złożone funkcje poznawcze mogą powstawać w wyniku współdziałania wielu stosunkowo prostych elementów. Dla informatyków stwarzają nadzieję na opracowanie nowych zasad równoległego przetwarzania informacji, realnej alternatywy dla tradycyjnych, sekwencyjnych komputerów, wreszcie dla fizyków są klasą modeli pozwalających zrozumieć i modelować procesy samoorganizacji zachodzącej w systemach złożonych. Sieci te w odróżnieniu od klasycznych modeli systemów poznawczych, wywodzących się z badań nad sztuczną inteligencją, doskonale dają sobie radę z brakiem czy zniekształceniem informacji, potrafią na podstawie uprzednich doświadczeń rozwiązywać zadania, dla

których nie da się sformułować prostych algorytmów rozwiązań.

Modele omawiane w tym artykule, w odróżnieniu od konstruowanych przez neurofizjologów, są budowane na bardzo uproszczonych założeniach. Procesy generacji potencjałów czynnościowych czy ich przewodzenia są bardzo skomplikowanymi procesami fizyko-chemicznymi. Na pewnym poziomie ogólności dokładna charakterystyka tych procesów nie ma istotnego wpływu na modelowane procesy. Dla działania układu nerwowego, czy nawet jego podsieci, podstawowe znacznie ma matematyczna charakterystyka funkcjonowania neuronu, nie zaś szczegółowy opis mechanizmów odpowiedzialnych za jego działanie, jak na przykład mechanizm otwierania i zamykania kanałów jonowych. Głównym i podstawowym założeniem wszystkich modeli sieci neuronowych jest równoległe przetwarzanie informacji. Sieć składa się z wielu neuronów połączonych ze sobą synapsami. Stan każdego neuronu w danej chwili zależy od sygnałów, które dopłynęły doń od innych neuronów w chwili poprzedniej.

SIEĆ ATRAKTOROWA JAKO UKŁAD FIZYCZNY

Z punktu widzenia fizyki podstawową cechą łączącą prezentowane tu modele sieci neuronowych i mózg jest fakt, że oba te systemy są układami dynamicznymi i przemiany w nich występujące można rozpatrywać w tych kategoriach. Są to układy zmienne w czasie, w których stan uprzedni układu wyznacza stany następne. Dynamiczne własności układu, takie jak: synchronizacja jego elementów czy bardziej ogólnie — koherencja w takim układzie, zmiany jakościowe trajektorii charakteryzującej jego dynamikę, pętle sprzężenia zwrotnego oraz kontrola dynamiki przez zmianę parametrów układu odgrywają podstawowe znaczenie podczas wykonywania odpowiednich funkcji.

Nasuwa się więc pytanie, jaki poziom opisu jest potrzebny do charakteryzacji tych zjawisk, jakie zmienne dają wgląd w makroskopowe własności dynamiczne takiego układu. Ogólnej odpowiedzi na te pytania udziela fizyka statystyczna i teoria nieliniowych układów dynamicznych. W ramach tych nauk wprowadza się pewien język opisu, który umożliwi badanie dynamicznych zmian zachodzących w układzie. Okazuje się, tak jak już zostało to wspomniane wcześniej, że niektóre szczegóły strukturalne w takich układach, jak również w pewnym stopniu dynamika poszczególnych elementów są nieistotne przy budowie obrazu ogólnego funkcjonowania układu. Określone grupy zmiennych makroskopowych pozwalają na stosunkowo dokładne i znacznie bardziej przejrzyste opisanie własności dynamicznych układu.

Dany układ fizyczny jest opisywany przez trzy podstawowe grupy parametrów (SCHUSTER 1995). Wielkości ulegające szybkim zmianom opisują mikroskopowe zachowanie się układu w czasie. Te szybko zmieniające się zmienne, tworzące pierwszą grupę, pozwalają zwykle na charakteryzację chwilowych własności układu. W modelach sieci neuronowych odpowiadają one opisowi dynamiki poszczególnych neuronów. Druga grupa, zwana parametrami, definiuje interakcje układu z otoczeniem, bądź pewne jego makroskopowe własności, które charakteryzują jego strukturę. Nawet niewielkie zmiany wartości takich wielkości mogą w zasadniczy, jakościowy sposób zmienić dynamikę układu — wtedy nazywa się je parametrami kontroli układu. Takie jakościowe zmiany podstawowych własności układu pod wpływem nawet bardzo niewielkich zmian parametrów kontroli nazywa się przejściami fazowymi. Poszczególne fazy definiują to, co można by na-

zwać dynamiczną strukturą układu. Najprostszym przykładem przejść fazowych jest przykład przejścia wody w lód (przejście fazowe pierwszego rodzaju), czy przejście ferromagnetyku w paramagnetyk w układach spinowych (przejście fazowe drugiego rodzaju).

Własności poszczególnych faz są opisywane przez trzecią grupę zmiennych, które są nazywane parametrami porządku. Wielkości te charakteryzują pewne makroskopowe własności układu jako całości. W przypadku przejścia fazowego wody w lód parametrem porządku może być gęstość, natomiast w przypadku układów spinowych — magnetyzacja układu. Przykładem parametru kontroli w obu tych układach może być temperatura. Okazuje się, że pewne ważne własności są specyficzne dla poszczególnych typów przejść fazowych i nie zależą od układu fizycznego, w którym zachodzą. Przejście fazowe w układach magnetycznych ma z pewnego punktu widzenia, takie same własności jak bifurkacja (podwajanie okresu), zachodząca w procesie opisywanym funkcją zwaną „mapą logistyczną” (SCHUSTER 1995), która opisuje m.in. dynamikę populacji. Oczywiście w obu tych układach mamy do czynienia z innymi parametrami kontroli (odpowiednio temperatura i parametr r) i różnymi parametrami porządku (odpowiednio wcześniej wspomniana magnetyzacja i współczynnik Lapunowa) ale przykładowo, parametry porządku skalują się w identyczny sposób w pobliżu punktu krytycznego.

Właśnie fakt, że pewne makroskopowe własności układów są uniwersalne i nie zależą od detali ich struktury, pozwala fizykom poznawać własności skomplikowanych i trudnych do opisu układów przez analogię do znacznie prostszych modeli. Opisany poniżej przykład modelowania wpływu stanu emocjonalnego na procesy poznawcze ma właśnie taki charakter. Struktura tego modelu w znacznym stopniu odbiega od tego, co wiemy o układzie nerwowym. Prezentowany model pokazuje jednak, po pierwsze, jak parametr porządku sieci neuronowych pozwala nam wykryć nowość wzorca, po drugie, jak mechanizmy kontroli (w tym przypadku emocje) mogą wpływać na procesy poznawcze zmieniając wartości parametrów kontroli sieci neuronowej. Zmiany wartości parametrów kontroli prowadzą do jakościowo innej dynamiki procesów zachodzących w modelach sieci neuronowych.

BUDOWA SIECI ATRAKTOROWEJ

Omawiane tu modele sieci neuronowych są budowane w analogii do struktury układu nerwowego. Podstawowymi elementami modeli sieci neuronowych są: węzły sieci, nazywane przez analogię do odpowiadających im elementów rzeczywistej sieci neuronowej neuronami, oraz połączenia między elementami, nazywane przez analogię synapsami (por. HERTZ i współaut. 1991). Neuron w tych modelach jest stosunkowo prostym elementem działającym na zasadzie wejście-wyjście (McCULLOCH i PITTS 1943). Stan neuronu jest określany liczbą, która opisuje stopień jego pobudzenia. W najprostszych modelach przyjmuje się, że neuron może przyjmować tylko jeden z dwóch stanów: wysoko pobudzony (+1) lub nisko pobudzony (-1).

Zasadą funkcjonowania sieci jest to, że stan każdego neuronu zależy od sumy sygnałów przychodzących od innych neuronów, z którymi

jest on połączony. Cała informacja znajduje się właśnie w sile i znaku połączeń między neuronami. Połączenia są opisywane przez liczby; im większa wartość bezwzględna liczby, tym silniejsze połączenie (tzn. tym silniej pierwszy z neuronów wpływa na drugi). Połączenia o znaku dodatnim są połączeniami pobudzającymi, zaś o znaku ujemnym — połączeniami hamującymi. Wartość sygnału dochodzącego od neuronu j do neuronu i określa się mnożąc liczbę opisującą stan neuronu j (w najprostszych modelach +1 lub -1) przez liczbę opisującą wartość połączenia synaptycznego między neuronem j a neuronem i . Sygnały dochodzące od innych neuronów do danego neuronu sumują się. Jeżeli suma ta przekroczy pewną wartość progową, to neuron staje się wysoko pobudzony, jeżeli nie, to znajduje się w stanie niskiego pobudzenia.

ROZPOZNANIE — DYNAMIKA NEURONÓW

Percepcji w modelu takiej sieci odpowiada ustalenie stanów wszystkich neuronów w sieci. Sieć rozpoczyna następnie ewolucję w taki sposób, że poszczególne neurony dopasowują swój stan do sygnału dochodzącego po synapsach od innych neuronów. Po pewnym czasie sieć osiąga stan równowagi, w którym konfiguracja stanów neuronów jest uważana za odpowiedź sieci. Wpływ neuronu j na neuron i jest wyrażany przez iloczyn stanu neuronu j (+1 albo -1) i siły połączenia między neuronami j oraz i . Jeżeli między dwoma neuronami istnieje połączenie pobudzające to mają one tendencję do tego, by być w tym samym stanie; jeżeli istnieje połączenie hamujące, mają one tendencję do bycia w stanach przeciwnych. Każdy węzeł w następnym momencie czasu ($t + 1$) przyjmuje stan określony przez sumę sygnałów dochodzących od innych neuronów w czasie t .

$$h_i(t) = \sum_{j \neq i} J_{ij} S_j(t) \quad (1)$$

W tym równaniu $h_i(t)$ określa całkowity wpływ wywierany na neuron i przez pozostałe neurony w czasie t , J_{ij} opisuje połączenia od neuronu j do neuronu i , a $S_j(t)$ opisuje stan neuronu j w czasie t . Jeżeli suma ta jest większa niż określona wartość progowa, w następnym momencie węzeł przyjmie stan wysoki, w przeciwnym razie przyjmie on stan niski. W przypadku gdy wartość progu jest równa 0, zależ-

ność pomiędzy sygnałem dochodzącym do neuronu w chwili t a stanem neuronu w chwili $t+1$ jest określona wzorem:

$$s_i(t+1) = \text{sgn}(h_i(t)) \quad (2)$$

gdzie:

$$\text{sgn}(x) = \begin{cases} 1 & \text{gdy } x > 0 \\ 0 & \text{gdy } x < 0 \end{cases} \quad (3)$$

Powyższe równania definiują dynamikę układu. W symulacjach numerycznych dynamikę taką zwykle realizuje się techniką Monte Carlo. Wybiera się losowo jeden neuron i na podstawie powyższych równań określa się, jaki stan powinien on przyjąć w następnej chwili, po czym wybiera się losowo następny.

W sieciach atraktorowych nie ma wyróżnionych warstw, charakteryzuje je powszechne występowanie pętli sprzężenia zwrotnego. Lokalna dynamika każdego z elementów układu jest skoordynowana z dynamiką innych elementów i wpływa na pewne globalne własności całego układu. Podstawową własnością sieci atraktorowych jest to, że sieć znajdująca się w dowolnym stanie podlega ewolucji w kierunku jednego ze stanów asymptotycznych, tak zwanych atraktorów. Stany te nazywają się atraktorami, gdyż układ do nich dąży. Wzorce, które sieć pamięta są właśnie atraktorami w takich sieciach.

Proces uczenia polega na takim ukształtowaniu połączeń w sieci, by wzorce, które mają

być zapamiętane stały się atraktorami w sieci. Proces rozpoznania zniekształconego lub niepełnego wzorca można przedstawić w sieci atraktorowej jako ewolucję w kierunku najbliższego atraktora i pozostanie w tym stanie przez pewien czas, niezbędny do pełnej identyfikacji bodźca. Doświadczenia MIYASHITA i CHANG (1988) pokazały, że w mózgu rzeczywiście występują atraktory. W doświadczeniach przeprowadzonych na małpach autorzy ci pokazali, że

aktywność poszczególnych neuronów po prezentacji wcześniej wyuczonego bodźca pozostawała nie zmieniona przez pewien czas (pomimo braku tego bodźca). Takie przedłużone pobudzenie świadczy o tym, że układ (w tym przypadku pewne neurony) został wprowadzony przez bodziec w jakiś metastabilny stan równowagi (atraktor), w którym pozostaje przez jakiś czas po wyłączeniu bodźca.

METAFORA ENERGETYCZNA

Dynamikę rozpoznania w sieci atraktorowej można łatwo zrozumieć przedstawiając ją w metaforze energetycznej. Fundamentalnym wkładem HOPFIELDA (1982) w rozwój tej klasy modeli było właśnie wprowadzenie pojęcia energii dla konfiguracji stanów sieci neuronalnej. Definicja ta jest analogiczna do definicji energii w układach szkieł spinowych, z których opisywane tu modele się wywodzą. Ewolucja sieci jest wyznaczona prostą zasadą opartą na minimalizacji energii w układzie. Sieć zmieni stan elementu wtedy i tylko wtedy, gdy każda taka zmiana spowoduje zmniejszenie energii w całym układzie — w przeciwnym razie stan elementu pozostaje nie zmieniony. W ten sposób sieć ewoluuje w kierunku stanu równowagi, który jest lokalnym minimum energii w przestrzeni stanów. Te stany równowagi odpowiadają właśnie wzorcom pamiętany przez sieć. Stąd też sieci atraktorowe bywają nazywane programowalnymi układami dynamicznymi. Przy określonych wartościach połączeń synaptycznych wartość energii dla danej konfiguracji sieci można zdefiniować jako:

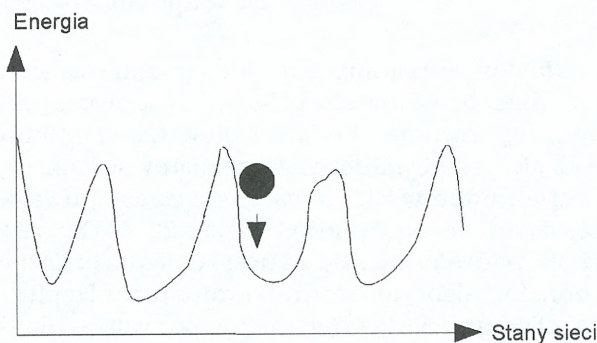
$$E = -\frac{1}{2} \sum_{j \neq i} J_{ij} S_i S_j \quad (3)$$

W tym równaniu J oraz S są używane w takim samym znaczeniu, jak używaliśmy ich w równaniu (1). Dynamikę stanów neuronów można dobrze przedstawić wprowadzając metaforyczny krajobraz energetyczny, w którym góry

odpowiadają stanom o wysokiej energii, doliny zaś stanom o niskiej energii.

Na osi odciętych są umieszczone konfiguracje stanów neuronów, tak że dwa sąsiednie punkty różnią się stanem jednego neuronu. Na osi rzędnych jest reprezentowana energia sieci.

W tym metaforycznym środowisku wzorcom



Rys. 1. Krajobraz energetyczny.

pamiętany przez sieć odpowiadają doliny — stany równowagi układu, czy też tak zwane atraktory — punkty przyciągające dynamiki. Ewolucję sieci w trakcie rozpoznawania można opisać jako piłkę toczącą się w górzystym krajobrazie w kierunku najbliższego dołka. Stanowi sieci, której pokazano zniekształcony wzór, odpowiada punkt na zboczu góry. Dynamika staczania się piłki odpowiada takim zmianom stanów neuronów w trakcie procesu rozpoznawania, że zniekształcenia są usuwane i jest odtwarzany zapamiętany wzorec.

DYNAMIKA W OBECNOŚCI SZUMU

Gdyby wpływy od innych neuronów były jedynym źródłem dynamiki, ustawałaby ona zawsze wtedy, gdy system osiąga stan minimum energetycznego. We wszelkich systemach

biologicznych zawsze jest obecny pewien szum. Ogólnie biorąc szum ten opisuje czynniki nie wynikające bezpośrednio z przekazywania informacji. W mózgu szum taki może być związa-

¹Pojęcie energii daje się wprowadzić tylko dla tych modeli, które zakładają symetryczność połączeń, tzn. zakładają, że połączenie od neuronu i do neuronu j jest takie samo, jak od neuronu j do neuronu i . Choć dla modeli z niesymetrycznymi połączeniami nie da się zdefiniować energii, ich ewolucja w kierunku atraktorów zachodzi w podobny sposób.

ny z wpływem aktywności innych ośrodków w mózgu, emisją rozmaitych ilości neurotransmitera przy danym pobudzeniu, czy nie zawsze taką samą amplitudą potencjału czynnościowego. Z pewnym przybliżeniem szum w takim układzie, tak jak w większości procesów fizykochemicznych, ma rozkład Gaussa. Szum o takim samym rozkładzie wprowadza się w opisywanych tu modelach.

Szum w modelach sztucznych sieci neuronowych wprowadza się jako dodatkowe, losowe oddziaływanie na neuron, określone prawdopodobieństwem tego, że neuron przyjmie stan niezgodny z wartością sygnału dochodzącego do niego. Prawdopodobieństwo to określa wzór:

$$p = \exp(-2h\beta);$$

gdzie β jest poziomem szumu w sieci, a h wartością sygnału dochodzącego do neuronów.

UCZENIE SIĘ — ZMIANY WARTOŚCI SYNAPS

Jak już wspomnieliśmy, ideą uczenia się sieci jest takie określenie siły połączeń synaptycznych, by stany równowagi odpowiadały wzorcom, których sieć ma się nauczyć. Dwie reguły uczenia się sieci atraktorowych są warte szczególnej uwagi. Algorytm perceptronowy (GARDNER i DERRIDA 1988) pozwala, by sieć nauczyła się dużej liczby wzorców. Algorytm sformułowany przez Hopfielda, choć nie jest tak optymalny, pozwala analitycznie określać różne właściwości sieci, dobrze więc się nadaje do budowy modeli.

Algorytm Hopfielda jest oparty na mającej podstawy neurofizjologiczne zasadzie sformułowanej przez HEBBA (1949), że jeżeli aktywność dwóch neuronów jest ze sobą silnie skorelowana, to siła połączenia synaptycznego między nimi będzie rosła. Jeżeli natomiast ich dynamika nie jest skorelowana, bądź ma korelację ujemną, to siła połączenia synaptycznego między tymi neuronami będzie słabła. Przekładając tę zasadę na język naszych modeli definiujemy algorytm, który będzie przypisywał wartość dodatnią danemu połączeniu, jeśli stan neuronów połączonych tą synapsą będzie zgodny (wysokie lub niskie pobudzenie) i ujemną w przeciwnym przypadku.

Szum jest tutaj parametrem kontroli układu i może w jakościowy sposób zmienić dynamikę sieci. W szczególności w obecności szumu neuron może przyjąć stan, który jest przeciwny (czyli wysoko aktywny lub niskoaktywny) do stanu wynikającego ze zsumowania dochodzących do niego wpływów z innych neuronów.

Jeżeli duża liczba neuronów zmieni swój stan, lub gdy dany stan jest mało stabilny (dołek jest płytki) w obecności szumu, sieć może opuścić stan równowagi. Stan ten traci stabilność. Im większy poziom szumu, tym bardziej stabilne atraktory tracą stabilność. Wreszcie przy pewnym, dużym poziomie szumu żaden ze stanów nie jest już stabilny, sieć nie jest w stanie odtworzyć żadnej z zakodowanych informacji, dynamika jest po prostu losowa.

Każdy wzorec może być reprezentowany jako wektor składający się z N dodatnich i ujemnych jedynek określających stany poszczególnych elementów sieci. Załóżmy, że sieć ma nauczyć się m takich wzorców. Jak wspomnieliśmy wyżej, celem uczenia jest sprawienie, by każdemu z tych wzorców odpowiadało lokalne minimum funkcji energii. Poniższy wzór wyraża siłę połączenia między neuronami i oraz j , jaką powinno ono przyjąć w wyniku uczenia:

$$J_{ij} = \frac{1}{N} \sum_{k=1}^m S_i^k S_j^k \quad (4)$$

gdzie J_{ij} opisuje siłę połączenia, S_i^k opisuje stan elementu i we wzorcu k , S_j^k opisuje stan elementu j we wzorcu k , m jest liczbą wzorców do nauczenia, a N liczbą neuronów w sieci. Jak wynika z tego wzoru, siła połączenia między dwoma neuronami jest proporcjonalna do wielkości korelacji między stanami, które przyjmują neurony kodując wzorce, których sieć ma się nauczyć.

STRUKTURA FAZOWA MODELU HOPFIELDA

W części poprzedniej omówiliśmy budowę i podstawowe założenia modelu Hopfielda. Okazuje się, że model Hopfielda nie zawsze jest w stanie uczyć się i rozpoznawać zadane wzorce. Z punktu widzenia fizyki, model Hopfielda jest zbudowany na bazie modelu szkieł spinowych i

ma z nim wiele wspólnych właściwości, w szczególności strukturę fazową. Poznanie struktury fazowej tego modelu pozwala zrozumieć jego ograniczenia co do jego możliwości rozpoznania. Co więcej, okazuje się, że sieci mogą rozpoznawać tylko w jednej z trzech możliwych faz

tego układu. Teraz więc zajmiemy się bliżej poznaniem struktury fazowej układu szkieł spinowych.

Podstawowymi parametrami, od których zależy faza układu, są, jak już wcześniej wspomniano, poziom szumu zwany również temperaturą w modelu energetycznym i pojemność układu. Zmiana tych parametrów może powodować jakościowe zmiany w dynamice układu. Granica pojemności sieci jest wyznaczona przez przejście fazowe. W wysokiej temperaturze lub przy wysokiej pojemności układ jest w tak zwanej fazie paramagnetycznej. W tej fazie neurony losowo zmieniają swój stan i układ nie jest w stanie osiąść w żadnym minimum. W metaforze krajobrazu energetycznego hiperpowierzchnia energii na osi stanów jest zupełnie płaska i system swobodnie po niej dryfuje.

Obniżając temperaturę w sposób ciągły (przejście fazowe drugiego rodzaju) wchodzimy w obszar fazy szkła spinowego. W tej fazie układ

jest sfrustrowany, gdyż nie jest jednak w stanie globalnie zminimalizować swojej energii, to jest posiada dużo lokalnych minimów. Układ więc ciągle wpada w różne minima w żaden sposób nie związane z nauczoną konfiguracją i zamara, (tzn. spowalnia dynamikę), dopóki fluktuacje termiczne nie wybiją go z tego stanu.

Trzecią fazą jest faza rozpoznawania nauczonych konfiguracji. W tej fazie nauczone wzorce stają się lokalnymi minimami energii i układ jest w stanie dokonać rozpoznania. Z tego krótkiego podsumowania struktury fazowej wiadać więc, jak różne może być zachowanie układu dla różnych wartości parametrów kontroli. Zmieniając jeden lub oba parametry kontroli możemy zmienić dynamikę układu z rozpoznania danego bodźca na losową zmianę stanów jego elementów. Wprowadzone poniżej pojęcie samokontroli opiera się właśnie na opisanym wyżej diagramie fazowym.

MODELE WPŁYWU EMOCJI NA PROCESY POSTRZEGANIA; SAMOKONTROLA TRAJEKTORII DYNAMICZNYCH UKŁADU

Modele przedstawione poniżej opierają się na tym, że atraktorowe sieci neuronowe są z punktu widzenia fizyki układami dynamicznymi. Wszystkie przedstawione modele są zbudowane według wspólnej zasady. Wykorzystują one fakt, że już w pierwszych chwilach dynamika sieci rozpoznającej nieznane wzorce jest inna od dynamiki sieci rozpoznającej wzorce znane. Sieć określa nowość wzorca zbierając informacje o swej własnej dynamice. Nowość wzorca wpływa na jeden z parametrów kontroli dyna-

miki: poziom szumu, jakościowo zmieniając przebieg procesu rozpoznania. Samokontrola układu polega na istnieniu pętli sprzężenia zwrotnego między parametrem porządku sieci neuronowej a parametrem kontroli tejże sieci. Sieć kontroluje więc przebieg własnej pracy, zmieniając wartości swych parametrów kontroli. Korzystając z tej zasady można zbudować różne szczegółowe modele, z których kilka przedstawimy poniżej.

JAK JEST WYKRYWANA NOWOŚĆ?

W poprzedniej części przedstawiliśmy wyniki badań doświadczalnych świadczące o tym, że ważną rolę w kształtowaniu reakcji emocjonalnej ma nowość bodźców. Co więcej, nowość bodźców jest wykrywana już w pierwszych chwilach procesu rozpoznania. W badaniach na modelach (LEWENSTEIN i NOWAK 1989a, b) wykazano, że nowość wzorca można określić na podstawie dynamiki sieci w pierwszych momentach procesu rozpoznania; sieć jest zdolna do błyskawicznego diagnozowania stopnia nowości bodźca na długo przed jego kategoryzacją. Zarówno symulacje komputerowe, jak i modele analityczne wykazały, że dynamika sieci wygląda zupełnie inaczej wtedy, gdy ma ona do czynienia z wzorcami nowymi, niż gdy podaje się jej bodźce znane. Różnica ta pojawia się już w pierwszych momentach przetwarzania informa-

cji, dużo wcześniej niż ostateczne rozpoznanie wzorca. Nowość bodźca można określić na podstawie dwóch różnych kryteriów, oba te kryteria są związane z właściwościami sieci jako układu dynamicznego.

Jak powiedzieliśmy wyżej, wzorcom zapamiętanym odpowiadają stany równowagi sieci. Stany równowagi — atraktory są określane poprzez konfiguracje stanów poszczególnych neuronów. Pierwsze kryterium jest związane bezpośrednio z bliskością atraktorów. Jeżeli sieć znajduje się w stanie równowagi, to jej dynamika jest bardzo powolna, konfiguracja stanów neuronów zmienia się wolno, a zmiany, które zachodzą, są spowodowane głównie szumem. Jeżeli sieć próbuje rozpoznać wzorec nieznany i nie podobny do już wyuczonych, to z punktu widzenia struktury istniejących połączeń sy-

naptycznych jest on wzorcem losowym. W sieci binarnej przeciętnie co drugi neuron będzie miał właściwą wartość ze względu na sygnały dochodzące od innych neuronów, a co drugi będzie chciał zmienić swą wartość. Dla wzorców trochę znanych szybkość zmian konfiguracji stanów neuronów jest wprost proporcjonalna do nowości wzorca.

Drugie kryterium różniące nowe wzorce od starych jest związane z siłą sygnałów dochodzących do neuronów. Sygnał dochodzący do danego elementu sieci jest wypadkową stanów wszystkich elementów układu, modulowanych poprzez wartości połączeń pomiędzy tym elementem a innymi. Okazuje się, że jeśli sieć jest w pobliżu stanu równowagi, to średnia wartość amplitudy sygnału, dochodzącego do danego neuronu, jest znacznie wyższa niż w przypadku gdy układ jest w jakimś losowym stanie nie związanym z żadną z nauczonych konfiguracji. Dzieje się tak dlatego, że w pobliżu stanu równowagi wszystkie elementy dążą do tego samego atraktora. Sygnały dochodzące do danego elementu od innych elementów wskazują na ten sam stan, sumują się. Gdy sieć znajduje się w stanie losowym, odpowiadającym percepcji nieznanego wzoru, różne fragmenty sieci będą dążyć do różnych atraktorów. Sygnały dochodzące do danego neuronu od innych neuronów będą konfliktowe, będą się nawzajem znosić, ich wypadkowa będzie mała. Kryterium to jest zresztą czulsze od poprzedniego i pozwala na dokładniejszą diagnozę stopnia nowości.

Niezależnie od tego, które kryterium jest używane, mechanizm wczesnej diagnozy nowości jest taki sam, sieć monitoruje swą dynamikę w pierwszych momentach rozpoznania i na tej podstawie może określić nowość, zanim właściwy proces rozpoznania dobiegnie końca. Takie

zachowanie sieci modeluje całą klasę fundamentalnych, a nie wyjaśnionych zjawisk psychologicznych, których istotą jest reagowanie organizmu na nowość bodźca przed świadomym jego rozpoznaniem. Wydaje się, że ta własność modelu może mieć bardzo istotne odniesienie do wyników eksperymentów neurofizjologów. Różne eksperymenty sugerują, że czasowe zależności pomiędzy potencjałami czynnościowymi poszczególnych neuronów w mózgu mogą mieć decydujące znaczenie w procesach rozpoznania. Synchronizacja czy koherencja sumarycznej aktywności grupy neuronów może być mechanizmem łączącym poszczególne procesy poznawcze na poziomie grup neuronów. Wzrastająca wartość amplitudy sygnału w pobliżu stanów równowagi może być uważana za pewnego typu koherencję sygnału w układach biologicznych.

Tak więc system poznawczy monitorując swą dynamikę może już w pierwszych momentach określić stopień nowości spostrzeganego wzorca. Przeprowadzono symulacje eksperymentów (SEAMON i współaut. 1984), polegających na porównaniu reakcji czysto afektywnej na nowość bodźca z rozpoznaniem (DROGOSZ i NOWAK 1996). Głównym wynikiem tych eksperymentów jest to, że przy bardzo krótkich czasach ekspozycji bodźca osoby badane nie są w stanie rozpoznać go, niemniej jednak reagują nań pozytywnymi emocjami, czyli system afektywny potrafi w jakiś sposób ustalić, że jest on znany. Przy dłuższych czasach ekspozycji rozpoznanie szybko się polepsza, zaś lubienie nie wzrasta już dalej. Symulacje zarówno z wykorzystaniem dynamiki zmian neuronów (DROGOSZ i NOWAK 1996), jak i amplitudy sygnału (NOWAK i współaut. 1997) odtworzyły te wyniki.

SAMOKONTROLA POZIOMU SZUMU

Emocje, jak już wspomnieliśmy, są czynnikami kontroli procesów poznawczych. W szczególności wzorce nieznane są przetwarzane w procesie rozpoznania inaczej niż wzorce znane. Prawidłowość ta daje się dobrze ująć w modelach atraktorowych sieci neuronowych. W modelach samokontroli spostrzeżona nowość wzorca może łatwo zostać wykorzystana do modyfikacji procesu postrzegania. W naszych modelach kontrolowaliśmy siłę, z jaką są pamiętane wzorce (LEWENSTEIN i NOWAK 1989a, b) przez wprowadzenie parametru wagi

$$J_{ij} = W^k \sum_{k=1}^m s_i^k s_j^k \quad (5)$$

gdzie W^k oznacza wagę, czyli siłę pamiętania wzoru k . Wysoka waga odpowiada silnie pamiętanym wzorcom, niska waga wzorcom ledwo pamiętanym.

Wpływ czynników afektywnych na procesy poznawcze modelujemy przez mechanizm, w którym stwierdzony stopień nowości decyduje o poziomie szumu w sieci. Bodźce nowe prowadzą do wzrostu poziomu szumu, co uniemożliwia ich rozpoznanie i stanowi odpowiedź „nie wiem”. Dla wzorców znanych lecz zniekształconych występuje średni poziom szumu, co prowadzi do mylnego rozpoznawania ich jako któregoś z bodźców silnie pamiętanych.

Modelowane przez nas mechanizmy samokontroli są mechanizmami niskiego poziomu, polegają głównie na istnieniu sprzężeń zwrotnych, gdzie stan sieci bądź parametry przebiegu procesu wpływają na właściwości sieci. Te mechanizmy samokontroli można interpretować jako wpływ czynników afektywnych na przetwarzanie informacji. Należy zwrócić uwagę, że zakładany mechanizm generowania odpowiedzi „nie wiem” jest zgodny z wynikami badań neurofizjologicznych uzyskanymi przez SKARDE i FREEMANA (1987), pokazującymi, że bodźce nieznanne prowadzą do chaotycznej dynamiki zespołów neuronów.

Zasady samokontroli w sieci można sformułować również w oparciu o lokalnie definiowane mechanizmy kontroli (ŻOCHOWSKI i współaut. 1994). W modelu tym każdy neuron określa siłę sygnału docierającego do niego i w zależności od niej zmienia swoją wartość szumu, czyli losowy składnik własnej dynamiki. Neuron dostosowuje też w pewnym stopniu swą wartość szumu do wartości szumu innych sąsiadujących neuronów. Sieć z lokalnie wbudowanymi mechanizmami samokontroli jest w stanie rozpoznać, które fragmenty bodźca są nowe (ewentualnie zniekształcone) i dokonać ich korekty.

W tradycyjnych modelach sieci neuronowych procesy uczenia są rozdzielone w czasie od procesów rozpoznawania. W sieci zmieniają się albo stany neuronów (proces rozpoznawania) albo wagi synaps (proces uczenia się). Takie rozdzielenie w czasie tych dwóch typów dynamiki nie jest realistyczne neurofizjologicznie. W mózgu bowiem procesy rozpoznawania zachodzą jednocześnie z procesami uczenia się. W modelach atraktorowych sieci neuronowych

zachodzą jednocześnie z procesami uczenia się. W modelach atraktorowych sieci neuronowych trudno jest połączyć te dwa typy dynamiki. Jeżeli zmieniają się stany neuronów w trakcie procesu rozpoznania, to jest zniekształcana reprezentacja wzoru, który miałby być zapamiętany; jeżeli zmieniają się wagi synaps podczas procesu uczenia się, to zmienia się również charakterystyka połączeń sieci. W rezultacie sieci, w których jednocześnie zachodzi zarówno zmiana wag synaps, jak i stanów neuronów są tak nieefektywne, że praktycznie nie potrafią funkcjonować.

Model samokontroli przedstawiony powyżej pozwala obejść wspomniane wyżej ograniczenie. Sieć SMARTNET (ŻOCHOWSKI i współaut. 1995), tak jak inne przedstawione tu modele, monitoruje nowość nadchodzących informacji. W sieci tej wprowadzono parametr, który określa stosunek dynamiki synaps do dynamiki neuronów. Wartość tego parametru zależy od nowości spostrzeganego wzorca. Natomiast dla wzorców znanych dominuje proces zmian stanu neuronów, a więc dynamika rozpoznania; zmiana synaps jest tak mała, że praktycznie jest zaniedbywalna. Im bardziej nie jest znany spostrzegany wzorec, tym słabsza jest zmiana stanów neuronów a silniejsza zmiana synaps. SMARTNET rozpoznaje więc wzorce znane, a uczy się wzorców nie znanych. Mimo połączenia dynamiki rozpoznawania z dynamiką uczenia się pojemność tego modelu jest równie duża jak modeli, w których uczenie się jest rozdzielone od rozpoznawania.

MODELOWANIE TREŚCI EMOCJI

Powyższe modele są modelami bardzo prostymi, w których modelujemy jedynie składnik pobudzenia występujący w emocjach. W oparciu o zaproponowaną metodę samokontroli można również modelować treść emocji (ŻOCHOWSKI i współaut. 1993). Jak wspomnieliśmy na wstępie, emocję oprócz stopnia pobudzenia układu oraz znaku (pozytywny–negatywny) można charakteryzować ze względu na jej treść, na przykład: smutek, gniew, radość lub zaskoczenie. Z badań psychologicznych wiadomo o występowaniu zgodności między treścią emocji a procesami poznawczymi. I tak, łatwiej przypominamy sobie materiał związany z treścią emocji, którą aktualnie przeżywamy, czyli na przykład, jeżeli jesteśmy smutni, to najłatwiej przypominamy sobie smutne wydarzenia. W następnym modelu pokazujemy, jak sieć może rozpoznać treść emocji związanej ze spostrzega-

nym bodźcem, zanim ten bodziec rozpozna. Informacja o znaczeniu emocjonalnym bodźca pozwala usprawnić proces rozpoznawania.

W modelu tym sieć jest podzielona na podsieci. Każda z podsieci koduje materiał związany z określoną emocją. Zasada działania systemu jest taka, że w jednej sieci kodujemy, na przykład, pamięć smutnych wydarzeń, w innej wesołych, w innej zaś gniewnych. Taka architektura systemu pozwala wykorzystać mechanizm szybkiego rozpoznawania nowości do rozpoznawania treści emocji związanej z rozpoznawanym wzorcem. Już w pierwszych momentach po spostrzeżeniu bodźca każda z podsieci decyduje niezależnie od innych, czy ten bodziec jest dla niej nowy, czy znany. Te sieci, które nie znają bodźca, stopniowo zwiększają poziom swego szumu, dzięki czemu po pewnym czasie wyłączają się z udziału w rozpoznaniu. Ta sieć,

która zna bodziec, zmniejsza swój poziom szumu. W konsekwencji rozpoznanie odbywa się wyłącznie w tej sieci, która zna bodziec.

Model najpierw więc rozpoznaje, jaka jest wartość emocjonalna bodźca, dopiero potem rozpoznaje sam bodziec w sieci kodującej bodźce, odpowiadające danej emocji. Symulacje komputerowe (ŻOCHOWSKI i współaut. 1993) wykazały, że taki, oparty o afekt, sposób kodowania jest znacznie efektywniejszy niż modele tradycyjne; w systemie o tej samej wielkości można zapamiętać kilkakrotnie więcej materiału! Użyte wyniki pozwalają więc sądzić, że procesy o charakterze emocjonalnym nie są po prostu przeżytkiem z poprzednich etapów ewolucji i przeszkodą w racjonalnym myśleniu, lecz wręcz przeciwnie, dzięki nim umysł jest w stanie sprawnie funkcjonować.

MODELOWANIE EMOCJI NEGATYWNYCH

Modele sieci z samokontrolą pozwalają również uwzględnić specyfikę emocji negatywnych, znaczy to, że są związane ze stanami, których organizm unika. W obszarze negatywnych reakcji emocjonalnych istnieje zjawisko analogiczne do omawianego wcześniej przed-rozpoznania, nazywane obronnością percepcyjną. Polega ono na tym, że selektywnie unika się rozpoznania wzorców, które skojarzone są z lękiem (na przykład w niektórych kulturach słów o treści seksualnej). Znowu pojawia się pytanie, skąd wiadomo, że powinno się unikać rozpoznania spostrzeganego wzorca, zanim nie rozpozna się jego znaczenia?

W celu modelowania zjawisk związanych z emocjami negatywnymi zdefiniowaliśmy nową klasę modeli sieci neuronowych, którą nazwaliśmy sieciami odpychającymi (NOWAK i współaut. 1993). Tradycyjnie w sieciach neuronowych stany odpowiadające zapamiętanym wzorcom definiuje się jako punkty przyciągające dynamikę — atraktory. Dynamika rozpoznania polega na tym, że sieć neuronowa startując z dowolnej konfiguracji zbliża się w trakcie rozpoznania do jednego z atraktorów odpowiadają-

Model ten pozwala zrozumieć w jaki sposób krótkotrwała, podprogowa prezentacja bodźca o wyraźnej wartości emocjonalnej sprawia, że ta wartość jest przypisywana bodźcom następnym (NOWAK i współaut. 1997). Zjawisko to jest nazywane w psychologii torowaniem. I tak, gdy prezentowane jest podprogowo zdjęcie uśmiechniętej twarzy, to następny bodziec zostanie oceniony jako bardziej pozytywny (MURPHY i współaut. 1995). Przyjmując interpretację zgodną z założeniami modelu można uważać, że dzieje się tak, gdyż prezentowany podprogowo bodziec zmniejsza szum w sieci kodującej wzorce pozytywne i zwiększa przez to szanse kodowania następnego bodźca właśnie w podsieci przeznaczonej do kodowania wzorców o pozytywnych skojarzeniach.

czego zapamiętanemu wzorowi. Dynamika taka jest więc dynamiką przyciągania. Sieci repellerowe są rządzone regułami odpychania. Sелеktywnie unikają one wybranych wzorców, rozpoznają wszystkie wzorce poza wzorcami zakazanymi. Zbliżanie się do jednego ze stanów, określonych jako odpychający, powoduje, że neurony wysyłają spójny sygnał do innych neuronów, by zmieniły swój stan. Taki mechanizm unikania może leżeć u podłoża negatywnej reakcji emocjonalnej. Sieci te mogą wyjaśniać takie zjawiska psychologiczne, jak tendencja do unikania, obronność percepcyjna, wyparcie czy niektóre mechanizmy uwagi.

W ramach jednej sieci repellery można łączyć z atraktorami. Sieci takie przejawiają bardzo bogatą dynamikę. Sieci atraktorowo-repellerowe mogą znaleźć zastosowanie w budowaniu systemów, które potrafią funkcjonować przy nieprecyzyjnie wyznaczonych zadaniach i niepełnej informacji. Mechanizmy samokontroli w analogii do funkcji emocji u człowieka mogą pomóc sieci orientować się w sytuacji niejasnych zadań i braku szczegółowych informacji o pożądanym celu zadania.

PODSUMOWANIE

Z pewnej perspektywy mózg jest po prostu systemem przetwarzającym informacje. W tradycyjnej psychologii poznawczej, która tę perspektywę przyjmuje, funkcje mózgu są opisywane jako ciąg operacji na symbolach. Z takiej perspektywy pewne procesy zachodzące w mózgu opisuje się łatwo. Próba opisu w tych kategoriach innych funkcji, na przykład reakcji

emocjonalnych i ich wpływu na procesy rozpoznania, może być niewygodna lub wręcz prowadzić do paradoksów.

Mózg, z innej perspektywy, jest również układem dynamicznym i wykazuje pewne ważne właściwości wspólne z innymi układami fizycznymi. Mózg ludzki jest prawdopodobnie najbardziej złożonym układem istniejącym w

znanym nam wszechświecie (por. EDELMAN 1992). Zrozumienie dynamiki mózgu jest ogromnym wyzwaniem pod adresem nauki i chyba nieprędko nauka w zadawalający sposób na to wyzwanie odpowie. Atraktorowe sieci neuronowe są również układami dynamicznymi. Mogą one być traktowane pod pewnymi względami jako ogromnie uproszczony model procesów zachodzących w mózgu. Zrozumienie dynamiki sieci i zasad tą dynamiką rządzących może przybliżyć w pewnym stopniu nasze rozumienie ogólnych zasad, na jakich zachodzą procesy przetwarzania informacji w mózgu.

W niniejszym artykule staraliśmy się przedstawić pewne bardzo proste zasady analizy funkcji atraktorowej sieci neuronowej jako układu fizycznego. Omówione tu modele sieci, a zwłaszcza model Hopfielda nie tylko nie jest realistycznym modelem rzeczywistego układu nerwowego, lecz pewne jego założenia, na przykład symetrii połączeń, wręcz nie są w układach biologicznych spełnione. Dzięki swej elegancji formalnej, wynikającej głównie ze zdefiniowania w tym modelu funkcji energii, model ten można traktować jako model wyjaśniania jakościowego procesów zachodzących w sieciach. W fizyce tego typu bardzo uproszczone modele stały się podstawą zrozumienia natury wielu zjawisk, na przykład z zakresu magnetyzmu, hydrodynamiki czy optyki. W oparciu o proste modele sztucznych sieci neuronowych można próbować więc zrozumieć procesy zachodzące w mózgu. Należy sobie zdawać oczywiście sprawę z uproszczeń zawartych w modelu i z tego że wyniki symulacji komputerowych mają głównie rolę heurystyczną.

Sieci neuronowe z mechanizmami samokontroli poziomu szumu wykazują wiele analogii do znanych mechanizmów wpływu procesów emocjonalnych na procesy poznawcze. Wyniki symulacji komputerowych modelu wykazują duże podobieństwo do wyników klasycznych eksperymentów przeprowadzanych na ludziach. Modele sieci z samokontrolą zależną od

nowości mogą zwiększyć nasze rozumienie wzajemnego wpływu procesów poznawczych i procesów emocjonalnych. Modele te wykazują większą elastyczność i większą plastyczność funkcjonowania oraz większą pojemność pamięci niż analogiczne modele sieci bez mechanizmów samokontroli.

Modele te oczywiście nie są realistycznymi modelami funkcjonowania układu nerwowego. Pokazują one jedynie pewne zasady, w oparciu o które może działać system nerwowy. Omówiona tu klasa modeli pokazuje ogólne zasady na jakich system poznawczy i system afektywny mogą nawzajem na siebie wpływać. Budowa tych modeli jest możliwa dzięki temu, że atraktorowe sieci neuronowe są programowalnymi układami dynamicznymi. Zarówno diagnoza nowości, jak i kontrola dynamiki są oparte na fizycznych właściwościach sieci. Emocje działają jako czynniki kontroli przebiegu procesów poznawczych. Modelowane tu mechanizmy, analogiczne do emocji, zwiększają plastyczność funkcjonowania sieci neuronowych, pozwalając na unikanie rozpoznawania bodźców nie znanych, a powodując uczenie się ich, unikanie bodźców awersyjnych i zwiększenie pojemności pamięci (ŻOCHOWSKI i wpółaut. 1993). Wyniki symulacji tych mechanizmów są zgodne z wynikami eksperymentów psychologicznych, dotyczących związku emocji z poznaniem.

Otwartym pytaniem jest, w jakim stopniu zaproponowane tu mechanizmy odpowiadają zagadnieniu, jak emocje wpływają na poznanie w realnych układach biologicznych, w rzeczywistych układach nerwowych. Niezależnie od tego, w jakim stopniu odpowiadają one rzeczywistym układom biologicznym, sieci z samokontrolą zbliżoną do mechanizmów wpływu emocji na poznanie mogą zostać wykorzystane do budowy sprawniejszych systemów sztucznej inteligencji. Samokontrola pozwala bowiem na szybką zgrubną kategoryzację i szybką reakcję, zwiększa plastyczność przetwarzania informacji oraz znacznie zwiększa pojemność pamięci.

AFFECT IN NEURAL NETWORKS

Summary

In this paper we propose a framework for modelling cognitive-affective interactions in artificial neural networks. The models presented are based on the results of psychological experiment and make use of the fact that attractor neural networks, from the point of view of physics, are programmable dynamical systems. Emotions are modelled as factors controlling cognitive processes by changing dynamical properties of the neural network. Introduction of

a self-control mechanism, by which a neural network controls its own dynamics, significantly improves performance of the network. The models suggest one of possible ways in which emotions can improve brain performance in the cognitive domain. Several models are presented which concern different properties of emotions and their influence on cognition.

LITERATURA

- CARVER C. S., SCHEIER M. F. 1990. *Origins and functions of positive and negative affect: A control-process view*. Psychological Review 97, 19-35.
- DAMASIO, A. R. 1994. *Descartes error. Emotion reason and the human brain*. New York: Putnam.
- DROGOSZ M., NOWAK A., 1996. EXAC, sieć neuronalna modelująca efekt ekspozycji. Przegląd Psychologiczny.
- EDELMAN G. M. 1992. *Bright air, brilliant fire*. New York: Basic Books.
- GARDNER E., DERRIDA B., (1988) *Optimal storage properties of neural network models*. Journal at Physics A 21, 271-284.
- HEBB D. O., 1949. *The Organization of Behavior: A Neurophysiological Theory*. New York: Wiley.
- HERTZ J., KROGH A., PALMER G. 1991. *Introduction to theory of neural computation*, Addison-Wesley Publishing, Redwood City.
- HOPFIELD J. J., 1982. *Neural networks and physical systems with emergent collective computational capabilities*. Proceedings of National Academy of Sciences USA 79, 2554-8.
- KUNST-WILSON W. R., ZAJONC R. B., 1980 *Affective discrimination of stimuli that cannot be recognized*. Science 207, 557-558.
- LAZARUS R. S., 1982. *Thoughts on the relation between emotion and cognition*. American Psychologist 37, 1019-1024.
- LEWENSTEIN M., NOWAK A. 1989a. *Fully connected neural networks with self-control of noise levels*. Physical Review Letters 62, 225-228.
- LEWENSTEIN M., NOWAK A., 1989b. *Recognition with self-control in neural networks*. Physical Review A40, 4652-64.
- MCCULLOCH W. S., PITTS W. A., 1943. *A logical calculus of the ideas immanent in neural nets*. Bulletin of Mathematical Biophysics 5 p. 115-124.
- MIYASHITA Y., CHANG H. S., 1988. *Neuronal correlate of pictorial short-term memory in the primate temporal cortex*. Nature 331, 68-70.
- MURPHY S. T., MONAHAN J.L., ZAJONC R. B., 1995. *Additivity of nonconscious affect: combined effects of priming and exposure*. Journal of Personality and Social Psychology 69, 589-602.
- MURPHY S. T., ZAJONC R. B., 1993. *Affect, cognition and awareness: Affective priming with suboptimal and optimal stimulus*. Journal of Personality and Social Psychology 69, 589-602.
- NOWAK A., LEWENSTEIN M., TARKOWSKI, W. 1993. *Repeller neural networks*. Physical Review E 48, 1491-1498.
- NOWAK A., ZAJONC R. B., LEWENSTEIN M., ŻOCHOWSKI M., 1997. *Affnet: A network models of cognitive-affective interaction*. Nie publikowany manuskrypt.
- OATLEY K., JOHNSON-LAIRD P. N., 1987. *Toward a cognitive theory of emotions*. Cognition and Emotion 1 29-50.
- PLUTCHIK R., 1980. *Emotion: A psychoevolutionary synthesis*. New York, Harper i Row.
- PLUTCHIK R. 1962. *The emotions: Facts theories and a new model*. New York: Random House.
- POSNER M., 1981. *Chronometric explorations of mind*. New York: Plenum.
- RUMELHART D. E., MCCLELLAND J. L. (red.) 1986. *Parallel Distributed Processes: Explorations in Microstructure of Cognition*. t. I. Cambridge, MA: MIT Press.
- REYKOWSKI J., 1972. *Eksperymentalna psychologia emocji*. Warszawa, PWN.
- SEAMON J. J., MARSH, R. L. BROODY N., 1984. *Critical importance of exposure duration for affective discrimination of stimuli that are not recognized*. Journal of Experimental Psychology: Learning Memory and Cognition 10, 465-469.
- SHUSTER G. H., 1995. *Chaos deterministyczny: Wprowadzenie*. Warszawa: PWN.
- SIMON H. A., 1967. *Motivational and emotional controls of cognition*. Psychological Review 74, 29-39.
- SKARDA C. A., FREEMAN W. J., 1987. *How brains make chaos to make sense of the world*. Behavioral and Brain Sciences 10, 161-195.
- SOKOLOV E. N., 1963. *Perception and the conditioned reflex*. London: Pergamon Press.
- SOKOLOV E. N., 1969. *The modeling properties of the nervous system*. [W:] M. COLE, I. MALTZMAN (red.) *A handbook of contemporary Soviet psychology*. New York: Basic Books, 1969.
- ZAJONC R. B., 1980. *Feeling and thinking: Preferences need no inferences*. American Psychologist 35, 151-175.
- ZAJONC R. B., 1997. *Emotions*. [W:] D. GILBERT, S. FISK, G., A., LINDZEY (red) *Handbook of Social Psychology*. New York: Guilford
- ŻOCHOWSKI M., LEWENSTEIN M., NOWAK A., 1994. *Local noise in neural networks with self-control*. International Journal of Neural Systems 5, 287-298.
- ŻOCHOWSKI M., LEWENSTEIN M., NOWAK A., 1995. *SMARTNET: A neural network with self-controlled learning*. Network, 6, 93-101.
- ŻOCHOWSKI M. LEWENSTEIN M., NOWAK A., 1993. *Memory that tentatively forgets*. Journal of Physics A 26, 2453-2460.