Podstawy sztucznej inteligencji

wykład 6

Sztuczne sieci neuronowe (SSN)

Joanna Kołodziejczyk

11 stycznia 2011

Plan wykładu

Uczenie nadzorowanie sieci wielowarstwowej

Wagi

„Inteligencja" sztucznej sieci neuronowe kryje się w wartościach wag pomiędzy neuronami. Zatem konieczna jest metoda dostosowywania wag w celu rozwiązania danego problemu.

Wsteczna propagacja błędów

Dla sieci typu MLP (wielowarstwowy perceptron), najczęściej wykorzystuje się algorytm uczenia BP (backpropagation), czyli wstecznej propagacji błędów.

Uczenie nadzorowane

Sieć w tej metodzie uczy się przez przykłady, czyli, należy dostarczyć zestaw próbek składający się z par wejście-wyjście. Wyjście to znany poprawny wynik dla każdego przypadku. Tak więc znane jest oczekiwane zachowanie sieci neuronowej, a algorytm BP pozwala na takie właśnie dopasowanie.

Algorytm BP (wersja uproszczona

1. Ustaw wagi na wartości losowe z przedziału $[-1, 1]$, $k = 0$
2. Dopóki nie osiągnięto założonego progu błędu $e(L) = Err$ lub nie wykonano założonej liczby kroków $k < MaxEpoch$:
   1. Wylosuj ze zbioru próbek dowolny punkt i oblicz odpowiedź sieci , $i = L$
   2. Wykonuj dla warstw $i > 0$
      1. Aktualizuj wagi połączeń w warstwie i zgodnie ze wzorem

$$\mathbf{w}(k+1, i) = \mathbf{w}(k, i) + \eta \mathbf{e}(i)\mathbf{x}_i(i)$$

      2. $i = i - 1$
   3. $k = k + 1$

Gdzie $\mathbf{e}(i)$, to błędy popełniane przez neurony w warstwie (i). W warstwie wyjściowej błąd obliczany jest na podstawie różnicy sygnału oczekiwanego i generowanego przez sieć. W warstwach ukrytych obliczany jest z błędu warstwy następnej.

Słowny opis algorytmu BP

- W kolejnych iteracjach wykonuje się kroki: jedna z próbek uczących jest wprowadzana do sieci.
- Na jej podstawie w oparciu o aktualny stan wag (początkowo wyniki będą losowe) sieć będzie dawała odpowiedź.
- Odpowiedź ta jest porównywana do znanego wyjścia z próbki, obliczany jest średni błąd kwadratowy.
- Wartość błędu jest następnie propagowana wstecz przez sieć i wprowadzane są małe zmiany w wartościach wag w każdej warstwie.
- Zmiany wag wykonuje się tak, by zmniejszyć wartość błędu dla danej próbki.
- Cały proces jest powtarzany dla każdej próbki, i może być wykonywany wielokrotnie.
- Cykl ten powtarza się, aż całkowita wartość błędu spada poniżej progu, który jest z góry określony.

Ważne uwagi do BP

- Metoda działa tak, by minimalizować funkcję błędu $E = \frac{1}{2}\sum_{o=1}^{N_o}(y_o - output_o)$, gdzie y_o — oczekiwana wartość wyjścia a $output_o$ wartość obliczona przez sieć, a N_o — liczba wyjść.
- Podczas nauki, wagi każdego połączenia są zmieniane przez wartość, która jest proporcjonalna do sygnału błędu.
- Można powiedź, że sieć nauczy się rozwiązywania problemu „wystarczająco dobrze". Właściwie sieć nigdy nie będzie tworzyć dokładnego odwzorowania ($Err \neq 0$), ale będzie raczej asymptotycznie zbiegać do funkcji idealnej.
- Ważne, by funkcja aktywacji była funkcją różniczkowalną, gdyż minimalizacja funkcji błędu wiąże się z wyznaczeniem gradientu.

Współczynnik uczenia η

Znaczenie

Reguła uczenia wymaga, by zmiana wag była proporcjonalna do gradientu. Gradientowa reguła największego spadku oczekuje wykonywania nieskończonej liczby kroków. Stałą skalującą jest współczynnik uczenia.

Wartości

W praktyce wybiera się maksymalnie duży współczynnik uczenia, ale taki, który nie prowadzi do oscylacji.

Współczynnik momentum α

Wzór na poprawkę wagi z momentum

$$\mathbf{w}(k+1,i) = \mathbf{w}(k,i) + \underbrace{\eta \mathbf{e}(i)\mathbf{x}_i(i)}_{\Delta w(k+1,i)} + \alpha \Delta w(k,i)$$

Przesłanką do zastosowania współczynnika momentum jest to, że poprzednie zmiany w wagach powinny mieć wpływ na aktualny kierunek przesuwania się w przestrzeni wag. Zatem momentum zapobiega oscylacjom.
Czyli, gdy wagi zaczną przemieszczać się w określonym kierunku w przestrzeni wag, to będą dążyć do dalszego ruchu w tym samym kierunku. Momentum pomaga przeskoczyć przez lokalne minima.

Dwa tryby aktualizowania wag

Tryb onliine

Aktualizacja wag po każdym zaprezentowanym wzorcu.

Tryb wsadowy (batch)

Aktualizacja po prezentacji wszystkich wzorców.

Jeśli współczynnik uczenia jest niewielki, to różnica pomiędzy dwoma procedurami jest niewielka. Znaczne różnice można zaobserwować, gdy współczynnik uczenia jest duży, jako że algorytmu wstecznej propagacji błędów zakłada, że pochodne błędów są zsumowane po wszystkich wzorcach.

Generalizacja

Generalizacja

Uogólnienie, operacja myślowa, polegająca na wykryciu przez uczącego się cechy lub zasady wspólnej dla danej klasy przedmiotów lub zjawisk (psych.). Jest to najbardziej oczekiwana cecha sieci.

Przeuczenie

Sieci bardzo skomplikowane (duża liczba wag) tworzą bardziej złożony model i są przez to bardziej skłonne do nadmiernego dopasowania. Przeuczenie rozpoznaje się po tym, że błąd uczenia jest bardzo mały, a błąd testowania jest duży.

Niedouczenie

Sieci z niewielką liczną neuronów (wag) mogą być niewystarczająco silne do zamodelowania poszukiwanej funkcji odwzorowania wejść w wyjście. Niedouczenie daje duży błąd uczenia. Prawie zawsze większe sieci generują mniejszy błąd.

Plan wykładu

Kiedy nie stosować MLP

- Czy można zapisać schemat lub wzór, który dokładnie opisuje problem? Jeśli tak, to t stosować tradycyjne metody.
- Czy istnieje już rozwiązanie sprzętowe lub softwarowe, które rozwiązuję większość problemu? Jeśli tak, to czas poświęcony na aplikację SSN może nie być tego wart.
- Czy funkcje mają „ewoluować' w kierunku, który nie jest z góry określony? Jeśli tak, wykorzystaj algorytm genetyczny.
- Czy łatwo wygenerować/pozyskać znaczną liczbę próbek wejście-wyjście pokazujących poprawne działanie? Jeśli nie, to nie będzie można trenować SNN.
- Czy problem jest bardzo „dyskretny"? Czy poprawną odpowiedź można znaleźć metodą Look-Up Table? Look-Up Table jest znacznie prostszy i bardziej dokładny.
- Czy wymagane są precyzyjne wartości liczbowe na wyjściach? SSN nie jest korzystne, gdy należy podać dokładne odpowiedzi liczbowe.

Kiedy stosować MLP

- Dostępna jest duża liczba danych typu wejście-wyjście, ale nieznana jest relacja pomiędzy wejściem i wyjściem.
- Problem wydaje się mieć ogromną złożoność, ale wyraźnie istnieje rozwiązanie.
- Łatwo jest wygenerować przykłady właściwego zachowania/działania.
- Rozwiązanie problemu może ulec zmianie, w granicach podanych parametrów wejściowych i wyjściowych (tzn. dzisiaj $2 + 2 = 4$, ale w przyszłości może się okazać, że $2 + 2 = 3.8$).
- Wyjścia mogą być rozmyte lub nienumeryczne.

Zastosowania sieci MLP

Jednym z najbardziej typowych zastosowań SSN jest przy przetwarzaniu obrazu:

- rozpoznawanie pisma odręcznego
- dopasowanie fotografii twarzy do zdjęć z bazy danych
- kompresji obrazu przy minimalnej utracie jakości

Inne aplikacje:

- rozpoznawanie mowy
- analiza podpisu radarowego
- przewidywania rynku akcji

Zastosowania SSN przykłady

- Omijanie przeszkód — nauka przez przykłady
  `http://www.youtube.com/watch?v=xLPrB5O5E_A`
- RoboSight `http://www.rnisys.com/`
- Połączenie sieci neuronowej z algorytmem genetycznym
  `http://www.youtube.com/watch?v=99DOwLcbK18&feature=related`

Uwaga

Stosując MLP nie trzeba znać rozwiązania problemu. W tradycyjnych metodach trzeba rozumieć zależność pomiędzy wejściem i wyjściem. Z siecią postępujemy tak, że pokazujemy jakie jest poprawne wyjście dla pewnego zestawu danych wejściowych.

Po odpowiednim czasie uczenia sieć odkryje zależność pomiędzy wejściami i wyjściami. Czasami dobrze jest zaprezentować sieci przykłady, które są nieistotne i sieć nauczy się je ignorować.

Natomiast ominięcie istotnych danych przyczyni się do tego, że sieć zbuduje niewłaściwą zależność wejście-wyjście.

Plan wykładu

Miles per Gallon

Zazwyczaj problemy rozwiązywane siecią neuronową są związane ze zbiorem statystyk [1]. Celem będzie wykorzystanie pewnych danych by przewidywać inne. Rozważmy bazę danych samochodów. Zawiera ona następujące pola:

- Waga samochodu
- Pojemność silnika
- Liczba cylindrów
- Konie mechaniczne
- Hybrydowy lub benzynowy
- Mil na galon

[1] Przykład zastosowania sieci `http://www.heatonresearch.com/content/non-mathematical-introduction-using-neural-networks`

Miles per Gallon cd.

Cel

Zakłada się, że zostały zebrane dane dla wymienionych pól, zatem możliwe jest zbudowanie sieci neuronowej, która jest w stanie przewidzieć, wartość jednego pola, na podstawie wartości innych pól. W tym przykładzie omówiony zostanie przykład prognozowania zużycia mil-per-galon.

Normalizacja danych

Należy zdefiniować problem w odniesieniu do macierzy wejść z wartościami rzeczywistymi z przyporządkowaną macierzą wyjść zawierającą również wartości rzeczywiste. Jest jeden dodatkowy wymóg do spełnienia. Zakres liczbowy każdego elementu w macierzy powinien być zawarty między 0 a 1 lub -1 i 1. Przekształcenie takie nazywane jest normalizacją.

Miles per Gallon cd.

Architektura SSN

Jest w sumie sześć danych. Pięć z nich będzie użyte, by przewidywać szóstą. Zatem sieć będzie miała pięć wejść i jedno wyjście.

Sieć będzie skonstruowana następująco:

Input Neuron 1: Waga samochodu
Input Neuron 2: Pojemność silnika
Input Neuron 3: Liczba cylindrów
Input Neuron 4: Konie mechaniczne
Input Neuron 5: Hybrydowy lub benzynowy

Output Neuron 1: Mil na galon

Miles per Gallon cd.

Zakresy wartości dla danych:

Waga samochodu: 100-5000 lbs
Pojemność silnika: 0.1 to 10 litrów
Liczba cylindrów: 2-12
Konie mechaniczne: 1-1000
Hybrydowy lub benzynowy: true lub false
Mil na galon: 1-500

Przyjęte zakresy mogą być trochę za duże jak na współczesne samochody. Jednakże, pozwoli to na minimalne zmiany sieci neuronowej w przyszłości. Dobrze jest też nie mieć zbyt wiele danych na skrajnych końcach przedziałów.

Miles per Gallon cd.

Normalizacja do zakresu 0-1

Jak znormalizować wartość 2000 lbs? Waga samochodu mierzona w zakresie od 100 dla 2000 to 1900 (2000-100). Wielkość całego zakresu wagi to 4900 lbs (5000-100). Zatem procent wartości w zakresie 1900/4900 to 38%. Ta wartość zasili wejście sieci neuronowej. Wartości tak obliczane będą spełniały kryteria zakresu 0-1.

Wartości nominalne

Wejście „hybrydowy" ma wartość true/false. Do reprezentowania tych wartości używa się 1 dla hybrydowych i 0 dla benzynowych. Wartości prawda/fałsz sprowadza się do dwóch wartości.

Jak zaprezentować obraz?

Obrazy są popularnym źródłem informacji dla sieci neuronowych. Przedstawiona niżej metoda kodowani obrazu jest prosta ale często skuteczna.

Zakłada się, że obraz, ma rozmiar 300x300 pikseli. Dodatkowo obraz jest kolorowy. Należy zatem 90000 pikseli pomnożyć razy trzy kolory RGB, co daje 270.000. W tym przypadku powinno się przyjąć sieć z 270.000 neuronami wejściowymi. To jest zdecydowanie za dużo dla sztucznej sieci neuronowej.

Jak zaprezentować obraz? cd.

Wynik downsamplingu do rozmiaru 32x32.

Sieć neuronowa wymaga teraz 1024 wejścia, i zakłożono, że patrzy się tylko na intensywność każdego kwadratu. Sieci neuronowe widzi zatem w bieli i czerni.

Jak zaprezentować obraz? cd.

Poziom nasycenia to wartość od 0 do 255 (to normalny zakres wartości RGB).
By dokonać normalizacji wartość wejściową wystarczy podzielić przez 255, na przykład: intensywność określona liczbą 10 staje się wejściem o wartości 10/255 lub 0,039.

Jak zaprezentować obraz? cd.

Wyjścia z sieci

W prezentowanym przykładzie oczekuje się od neuronów wyjściowych komunikowania jaki obraz widzi sieć neuronowa. Zwykle rozwiązaniem jest utworzenie jednego neuronu wyjściowego dla każdego rodzaju rozpoznawanego przez SSN obrazu. Sieć zostanie nauczona, aby wystawić na wyjściu neuronu odpowiadającemu za dany obraz wartość 1,0 .

Sieć neuronowa do danych zmieniających się w czasie

Finansowe sieci neuronowe

Są bardzo popularną formą temporalnej sieci neuronowych. Temporalne sieci neuronowe to takie, które akceptują na wejściu wartości zmienne w czasie. Istnieją dwa sposoby prezentowania danych zmiennych w czasie w sieci neuronowej. Jedną z nich jest użycie okna wejścia i okna predykcji.

Rozważmy sieć neuronową do przewidywania rynku akcji. Dostępne są następujące informacje na temat cen akcji, które reprezentują kursu zamknięcia dla akcji w ciągu kilku dni.

Sieć neuronowa do danych zmieniających się w czasie cd.

Pierwszym krokiem jest normalizacja danych. W tym celu zamienia się każdą wartość na procent w stosunku do wartości t z poprzedniego dnia. Na przykład, dzień 2 da 0,04, bo przesunięcie z 45 dolarów do 47 to 4% wzrostu w stosunku do 45.

Dzień	Dane przed normalizacją	Dane po normalizacji
Day 1	$45	
Day 2	$47	0.04
Day 3	$48	0.02
Day 4	$40	-0.16
Day 5	$41	0.02
Day 6	$43	0.04
Day 7	$45	0.04
Day 8	$57	0.04
Day 9	$50	-0.12
Day 10	$41	-0.18

Sieć neuronowa do danych zmieniających się w czasie cd.

Cel

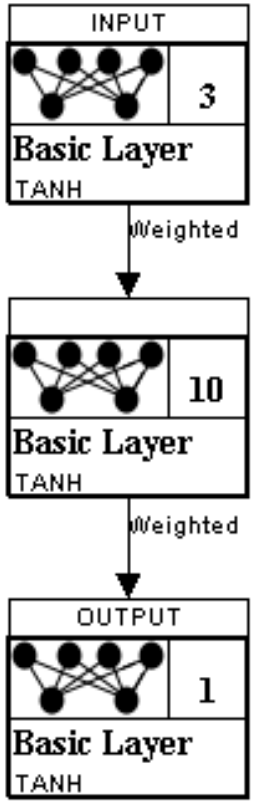
Utworzyć sieć neuronową, która będzie przewidywać wartości na następny dzień. Jak zakodować dane, które zostaną wprowadzone do sieci neuronowych. Sposób, w jaki to zrobimy zależy od tego, czy sieć neuronowa jest rekurencyjna.

Sieć neuronowa do danych zmieniających się w czasie cd.

Wielowarstwowy perceptron

Aby korzystać z tej sieci do przewidywania cen akcji musimy korzystać z „okien czasowych". Będziemy korzystać z okna trzech cen, by przewidzieć czwartą wartość. Tak więc, by przewidzieć jutrzejszy kurs patrzymy na ostatnie trzy dni. Oznacza to, że mamy trzy neurony wejściowe i jeden w warstwie wyjściowej.

Sieć neuronowa do danych zmieniających się w czasie cd.



Sieć neuronowa do danych zmieniających się w czasie cd.

Dane uczące dla wielowarstwowego perceptronu:

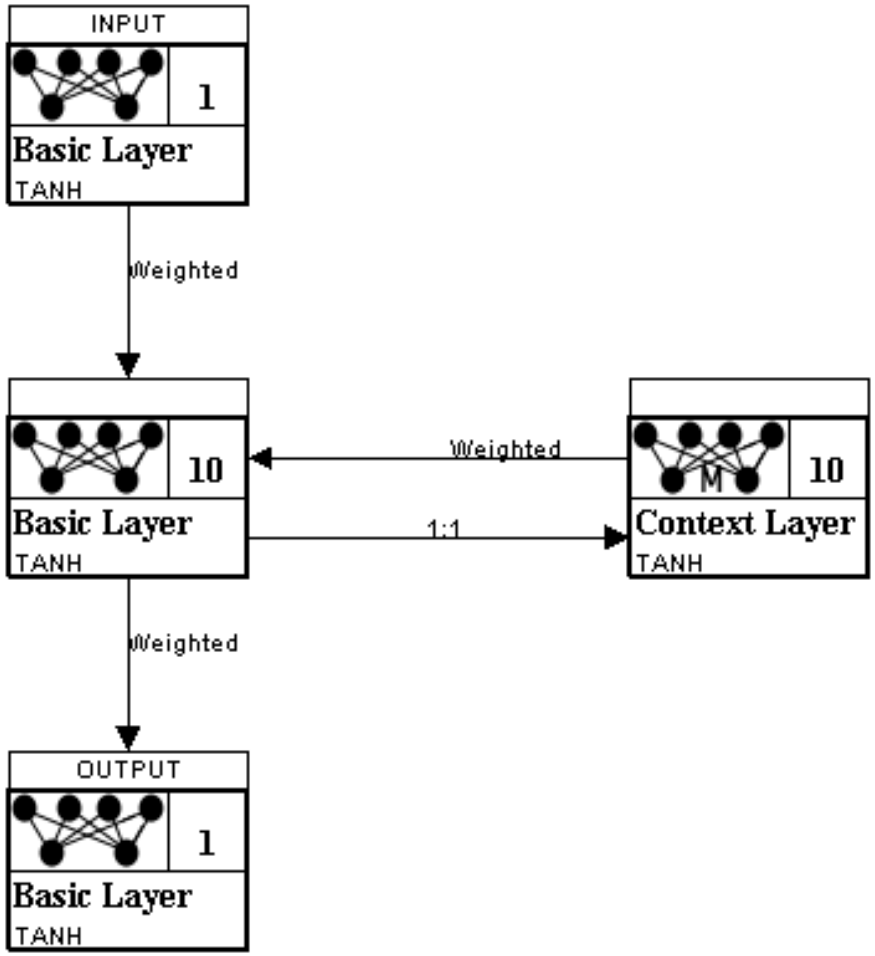
$(0.04, 0.02, -0.16) \rightarrow (0.02)$
$(0.02, -0.16, 0.02) \rightarrow (0.04)$
$(-0.16, 0.02, 0.04) \rightarrow (0.04)$
$(0.02, 0.04, 0.04) \rightarrow (0.04)$
$(0.04, 0.04, 0.04) \rightarrow (-0.12)$
$(0.04, 0.04, -0.12) \rightarrow (-0.18)$

Sieć neuronowa do danych zmieniających się w czasie cd.

Sieć rekurencyjna

Istnieje kilka różnych typów rekurencyjnych sieci neuronowych. Używana w tym przykładzie jest nazywana siecią neuronową Elmana. W sieci tej zastosowany zostanie jeden neuron wejściowy oraz pojedynczy neuron wyjściowy. Będzie można dostarczać tylko informacje z jednego dnia i oczekiwać zmiany kursy w dniu następnym.

Sieć neuronowa do danych zmieniających się w czasie cd.



Sieć neuronowa do danych zmieniających się w czasie cd.

Architektura sieci

Sieć neuronowa Elmana ma również 10 ukrytych neuronów. Tak jak było w przypadku sieci jednokierunkowych, wartość 10 została wybrana arbitralnie. W tej sieci funkcjonuje warstwa neuronów kontekstowa. Ta warstwa jest rekurencyjna. Warstwa kontekstowa pozwala zapamiętać pewien stan. To spowoduje, iż wyjście z sieci neuronowych zależało od więcej niż tylko ostatnie wejście.

Sieć neuronowa do danych zmieniających się w czasie cd.

Dane uczące dla sieci Elmana:

$(0.04) \rightarrow (0.02)$
$(0.02) \rightarrow (-0.16)$
$(-0.16) \rightarrow (0.02)$
$(0.02) \rightarrow (0.04)$
$(0.04) \rightarrow (0.04)$
$(0.04) \rightarrow (0.04)$
$(0.04) \rightarrow (-0.12)$
$(-0.12) \rightarrow (-0.18)$

Sieć neuronowa do danych zmieniających się w czasie cd.

Jak widać z powyższych danych, bardzo ważna jest kolejność prezentowanych próbek. Sieć neuronowa nie może przewidzieć przyszłej ceny widząc tylko cenę bieżącą. Dlatego, że warstwa kontekstowa pamiętająca ostatnie kilka dni będzie miała wpływ na wyjście. Można to porównać z rozpoznaniem piosenki. Trzeba usłyszeć przynajmniej kilka nut w sekwencji. Oczywiście kolejność dźwięków jest bardzo ważna.

Wprowadzanie tekstu do sieci neuronowej

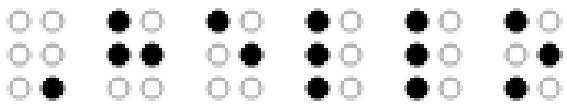
Wprowadzanie tekstu do sieci neuronowej jest wyjątkowo kłopotliwe z dwóch powodów:

1. słowo ma zmienną długość (sieć wymaga stałej liczby wejść i wyjść)
2. kodowanie znaków (a-z powinno być przechowywane w 1 czy 26 neuronach?)

Zastosowanie sieci Elmana rozwiązuje część z tych problemów. Stosując sieć Elmana z wystarczającą liczbą neuronów w warstwie wejściowej, by zapamiętać litery alfabetu i stosując warstwę kontekstową, by pamiętać kolejność liter można rozwiązać problem rozpoznawania tekstu. Analiza tekstu będzie się zatem wiązać z wprowadzaniem strumienia znaków.

Wprowadzanie tekstu do sieci neuronowej cd.

Kodowanie można rozwiązać stosując np. kody ASCII. Niniejszy przykład pokazuje zastosowanie znaków Braille'a, które łatwiej zakodować. Braille stosuje 6 kropek by reprezentować literę i znaki interpunkcyjne. Zakłada się, że te 6 kropek będzie wejściami do sieci neuronowej. Słowo „Hello" podane poniżej:



Pierwsze 6 kropek oznacza, że następny znak jest wielką literą. Zatem litera „h" (drugi znak od lewej) można zakodować jako (1, -1, 1, 1 -1, -1) (1 dla czarnej kropki, -1 dla białej).

Wprowadzanie tekstu do sieci neuronowej cd.

Cel

Sieć ma rozpoznać słowo „hello” i nie dać sygnału na wyjściu tak długo, jak długo wyraz się nie pojawi w całości.

Sieć Elmana będzie miała 6 wejść i 1 wyjście. Na wyjściu pojawi się 0 gdy wyraz nie rozpoznany i 1, gdy rozpoznano sekwencję znaków „hello”.

Dane uczące:

$[-1,-1,-1,-1,-1,1] \rightarrow [0]$
$[1,-1,1,1,-1-1] \rightarrow [0]$
$[1,-1,-1,1,-1-1] \rightarrow [0]$
$[1,-1,1,-1,1,-1] \rightarrow [0]$
$[1,-1,1,-1,1,-1] \rightarrow [0]$
$[1,-1,-1,1,1,-1] \rightarrow [1]$

Wprowadzanie tekstu do sieci neuronowej cd.

Prezentowany przykład pokazuje rozpoznanie pojedynczego słowa. Sieć Elmana powinna sobie radzić z wyrazami o zmiennej długości.

Aby rozpoznawać więcej wyrazów, trzeba dodać kolejny neuron wyjściowy. Rozpoznawanie języka naturalnego staje się więc przy takim systemie modelowania zbyt złożone.