Roman ZAJDEL
POLITECHNIKA RZESZOWSKA, KATEDRA INFORMATYKI I AUTOMATYKI

Uczenie ze wzmocnieniem regulatora Takagi-Sugeno metodą elementów ASE/ACE

Dr inż. Roman ZAJDEL

Studia na Wydziale Elektrycznym Politechniki Rzeszowskiej ukończył w 1990 roku, a stopień doktora uzyskał w 1999 roku na Politechnice Wrocławskiej. Obecnie pracuje na stanowisku adiunkta w Katedrze Informatyki i Automatyki PRz. Zajmuje się problematyką sieci neuronowych, systemów rozmytych oraz rozmytych sieci neuronowych ze szczególnym uwzględnieniem uczenia ze wzmocnieniem.

e-mail: rzajdel@prz-rzeszow.pl

Streszczenie

W artykule opisano zastosowanie algorytmu uczenia ze wzmocnieniem metodą elementów ASE/ACE do uczenia następników reguł regulatora rozmytego Takagi - Sugeno. Poprawność proponowanych rozwiązań zweryfikowano symulacyjnie w sterowaniu układem wahadło odwrócone - wózek. Przeprowadzono również eksperymenty porównawcze z klasyczną siecią elementów ASE/ACE. Pokazano zalety i wady rozwiązania klasycznego i rozmytego.

Abstract

The adaptation of reinforcement learning algorithm with the use of ASE/ACE elements for rule consequence learning of the Takagi - Sugeno fuzzy logic controller is proposed. The solution is applied to control of the cart-pole system and tested by computer simulations. The original neuronlike elements ASE/ACE are simulated as well. Advantages and disadvantages of the both approaches (fuzzy and classical) are demonstrated.

Słowa kluczowe: regulator rozmyty, uczenie ze wzmocnieniem, wahadło odwrócone.

Keywords: fuzzy controller, reinforcement learning, inverted pendulum.

1. Wstęp

Najczęściej stosowane algorytmy uczenia rozmytych systemów adaptacyjnych w większości ograniczają się do metody uczenia z nadzorem, która wymaga zestawu danych wejściowych i wzorcowych danych wyjściowych. Taki sposób pozyskania wiedzy ma tę wadę, że wiedza o procesie sterowania pochodzi wyłącznie od eksperta a regulator będzie tylko się uczył go naśladować. Ogranicza to możliwość uzyskania lepszych wyników sterowania.

Znacznie korzystniejszą wydaje się być metoda uczenia ze wzmocnieniem, nazywana często metodą uczenia z krytykiem. Nie wymaga ona określenia, jakie powinno być wyjście dla danego sygnału wejściowego, lecz ogranicza się tylko do krytyki, czy aktualnie podjęta przez system akcja przyniosła pozytywny, czy negatywny skutek.

Algorytm uczenia ze wzmocnieniem pierwotnie stosowano do elementów neuronowych [1,4], jednak w ostatnich latach daje się zauważyć zwiększone zainteresowanie połączeniem algorytmu uczenia ze wzmocnieniem z logiką rozmytą [2,3,5,7]. Niewątpliwą zaletą takiego rozwiązania jest możliwość pozyskania reguł sterowania w procesie uczenia opartym na minimalnej wiedzy eksperta. Ekspert jest zwolniony z konieczności „ręcznego" formułowania tablicy reguł, czyli formułowania ich przed procesem sterowania. Jest to o tyle istotne, że formułowanie bazy liczącej więcej niż kilkanaście reguł staje się procesem uciążliwym. Czy istnieją jeszcze inne korzyści takiego rozwiązania? Próbą udzielenia odpowiedzi na postawione pytanie jest niniejszy artykuł.

Na wstępie pokrótce opisano algorytm uczenia ze wzmocnieniem metodą elementów ASE (associative search element) i ACE (adaptive critic element) adaptowany do systemu rozmytego Takagi-Sugeno (T-S). Następnie przedstawiono wyniki eksperymentów polegających na sterowaniu wahadłem odwróconym umieszczonym na wózku za pomocą oryginalnego algorytmu ASE/ACE i jego rozmytej adaptacji. Na zakończenie przedstawiono krótkie podsumowanie.

2. Uczenie ze wzmocnieniem następników reguł regulatora Takagi - Sugeno

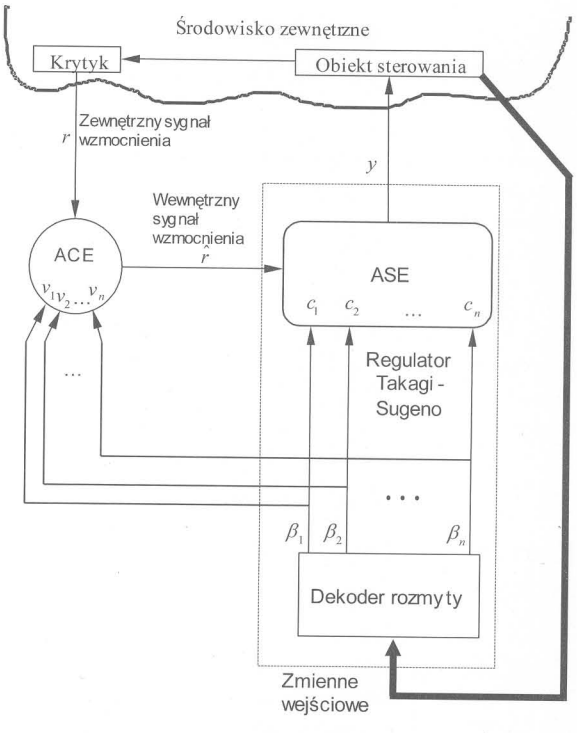
Algorytm uczenia ze wzmocnieniem elementów neuronowych ASE/ACE został zaproponowany przez Barto i Suttona [1]. Poniżej opisano adaptację tego algorytmu do doboru następników reguł c_n regulatora rozmytego T-S. Wyjście tego regulatora wyznacza się z zależności:

$$y = \frac{\sum_{n=1}^{N} \beta_n c_n}{\sum_{n=1}^{N} \beta_n} \qquad (1)$$

w której stopnie aktywności poszczególnych reguł dane są jako:

$$\beta_n(\mathbf{x}) = \mu_{A_{1n}}(x_1) \cdot \mu_{A_{2n}}(x_2) \cdot \ldots \cdot \mu_{A_{in}}(x_i) \cdot \ldots \cdot \mu_{A_{In}}(x_I) \qquad (2)$$

gdzie $\mu_{Ain}(x_i)$ jest stopniem przynależności zmiennej wejściowej x_i ($i = 1, ..., I$) do zbioru rozmytego A_{in}, zaś $n = 1,..., N$ jest indeksem reguły. Ideę uczenia następników reguł regulatora T-S przedstawiono na rysunku 1.



Rys. 1. Regulator rozmyty T-S uczony metodą elementów ASE/ACE
Fig. 1. Fuzzy controller T-S trained according to ASE/ACE elements method

Sieć składa się z dwóch elementów: regulatora rozmytego T-S (pełniącego funkcję skojarzeniowego elementu poszukującego ASE) i elementu neuronowego ACE. Środowisko zewnętrzne dostarcza systemowi wektor sygnałów wejściowych, na podstawie którego regulator T-S wytwarza sygnał wyjściowy y. Sygnał ten jest oceniany (w sensie jakości wykonania zadania sterowania) przez krytyka, przy czym ocena ta ma postać zewnętrznego sygnału wzmacniającego uczenie r. Element neuronowy ACE na podstawie sygnału r i rozmytej reprezentacji sygnałów wejściowych β_n wytwarza wewnętrzny sygnał wzmocnienia $\hat{r}$, będący podstawą algorytmu uczenia ze wzmocnieniem.

2.1. Dekoder rozmyty

Zadaniem rozmytego dekodera jest wyznaczenie stopnia aktywności poszczególnych reguł z zależności (2). Należy podkreślić istotną różnicę w funkcjonowaniu dekodera rozmytego w stosunku do dekodera dyskretnego z pracy [1]. Otóż, dekoder dyskretny klasyfikował, na podstawie zmiennych wejściowych, zaistniałą sytuację i uaktywniał jedno i tylko jedno swoje wyjście, podczas gdy dekoder rozmyty umożliwia, w zależności od typów wejściowych funkcji przynależności, uaktywnienie przynajmniej jednego wyjścia (β_n). Dodatkowo, wyjściem dekodera dyskretnego mogą być tylko dwa stany: 0 lub 1, podczas gdy dekoder rozmyty pozwala na płynne określenie (liczbami z przedziału [0,1]) przynależności danej sytuacji do jednej z możliwych kombinacji. Można więc na tej podstawie powiedzieć, że dekoder dyskretny jest szczególnym przypadkiem dekodera rozmytego.

2.2. Zmodyfikowany skojarzeniowy element poszukujący (ASE)

Zadanie elementu ASE polegające na poszukiwaniu optymalnego sterowania zostało przejęte przez regulator T-S, o wyjściu wyznaczanym z zależności [6]

$$y = \frac{\sum_{n=1}^{N} \beta_n c_n}{\sum_{n=1}^{N} \beta_n} + \delta(t) \qquad (3)$$

w której część deterministyczna wyjścia $\frac{\sum_{n=1}^{N} \beta_n c_n}{\sum_{n=1}^{N} \beta_n}$ zostaje wzbogacona częścią stochastyczną $\delta(t)$ w celu poszukiwania wyjścia najkorzystniejszego w sensie maksymalizacji całkowitej sumy wzmocnień. Wynik stochastycznego poszukiwania zostaje oceniony (skrytykowany) przez środowisko zewnętrzne. Ocena w formie sygnału wzmocnienia wpływa na parametry regulatora (następniki reguł c_n). Uczenie wartości c_n odbywa się zgodnie z zależnością

$$c_n(t+1) = c_n(t) + \alpha \hat{r}(t) e_n(t) \qquad (4)$$

w której α jest stałym dodatnim współczynnikiem uczenia, $\hat{r}(t)$ - wewnętrznym sygnałem wzmocnienia, $e_n(t)$ - uśrednionym parametrem dopasowania (eligibility) n-tego wejścia. Parametr ten (będący formą pamięci aktywności reguły i podjętej przez system akcji) również podlega uczeniu wg poniższej zależności

$$e_n(t+1) = \rho e_n(t) + (1-\rho) y(t) \beta_n(t) \qquad (5)$$

w której $0 \le \rho < 1$ jest stałym współczynnikiem.

2.3. Zmodyfikowany adaptacyjny element oceniający (ACE)

Podobnie jak w sieci klasycznej ASE/ACE zadaniem elementu ACE będzie wyznaczenie predykcji $p(t)$ sygnału wzmocnienia i na tej podstawie oszacowanie błędu predykcji (wewnętrznego sygnału wzmocnienia) $\hat{r}(t) = r(t) + \gamma p(t) - p(t\text{-}1)$. Stała γ określa względną ważność wzmocnień bliskich i odległych w czasie ($0 \le \gamma \le 1$). Sygnał predykcji został określony jako

$$p(t) = G\left(\sum_{n=1}^{N} v_n(t) \beta_n(t)\right) \qquad (6)$$

gdzie $G(\cdot)$ oznacza funkcję logistyczną określoną jako $G(x)=2/(1+e^{-\xi x})-1$. Reguła uczenia wag $v_n(t)$ ma natomiast postać

$$v_n(t+1) = v_n(t) + \chi \hat{r}(t) \bar{\beta}_n(t) \qquad (7)$$

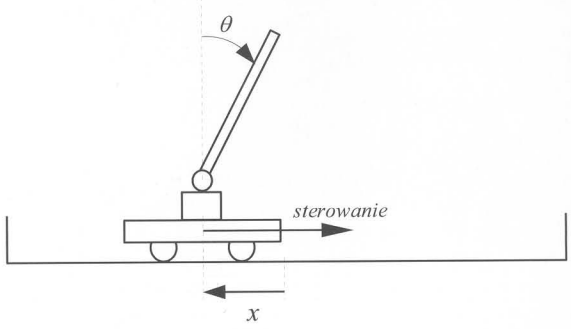
gdzie χ jest dodatnim współczynnikiem. Parametr $\bar{\beta}_n(t)$ jest śladem aktywności n-tej reguły rozmytej, określanym jako średnia ważona jej poprzedniego śladu i aktualnego poziomu aktywności

$$\bar{\beta}_n(t+1) = \lambda \bar{\beta}_n(t) + (1-\lambda) \beta_n(t) \qquad (8)$$

w której $0 \le \lambda < 1$ jest współczynnikiem określającym zanik śladu aktywności.

3. Eksperymenty

Eksperymenty polegały na zastosowaniu uczonego ze wzmocnieniem regulatora T-S (opisanego w poprzednim punkcie) do sterowania wahadłem odwróconym połączonym przegubowo z wózkiem (rys. 2). Dla celów porównawczych sterowanie przeprowadzono również przy pomocy oryginalnej sieci elementów ASE/ACE z pracy [1].



Rys. 2. Wahadło odwrócone umieszczone na wózku
Fig. 2. Cart-pole system

Eksperymenty zrealizowano w oparciu o podany w pracy [1] matematyczny model układu wahadło - wózek o czterech zmiennych stanu: [x (pozycja wózka), θ(odchylenie wahadła od poziomu), $\dot{x}$ (prędkość wózka), $\dot{\theta}$ (prędkość kątowa wahadła)].

Sterowanie jest wyznaczane na podstawie wyjścia y elementu ASE następująco:

$$sterowanie = \begin{cases} 10 & \text{jeśli } y > 0 \\ -10 & \text{jeśli } y \le 0 \end{cases}$$

W przypadku oryginalnej sieci elementów ASE/ACE wartości początkowe współczynników wagowych sieci akcji ASE (pełniącej rolę regulatora) są zerowe. W każdej dyskretnej chwili czasu t dekoder na podstawie wektora zmiennych stanu ustawia w stan wysoki jedno ze swoich wyjść, co jest równoznaczne z wyborem sterowania zapisanego w tabeli współczynników wagowych elementu ASE. Uniwersum poszczególnych zmiennych stanu podzielono na dyskretne przedziały następująco:

$$x \in \{[-2.4,-0.8] \cup [-0.8,0.8] \cup [0.8,2.4]\}\,[m]$$
$$\theta \in \{[-12,-6] \cup [-6,-1] \cup [-1,1] \cup [1,6] \cup [6,12]\}\,[^o]$$
$$\dot{x} \in \{[-\infty,-0.5] \cup [-0.5,0.5] \cup [0.5,\infty]\}\,[m/s]$$
$$\dot{\theta} \in \{[-\infty,-50] \cup [-50,50] \cup [50,\infty]\}\,[^o/s]$$

Liczba przedziałów 3 x 5 x 3 x 3 daje 135 możliwych stanów, (elementów czterowymiarowej tablicy sterowań).

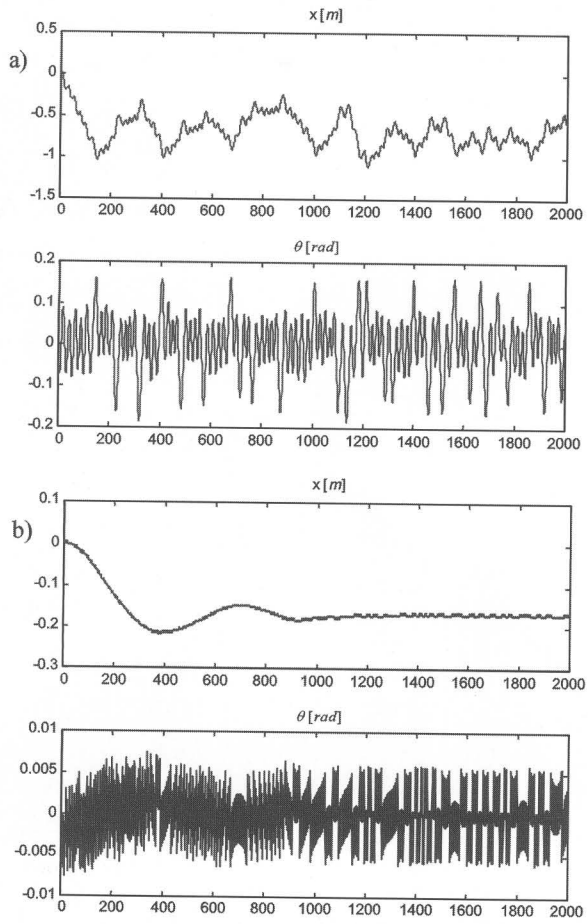
W przypadku regulatora T-S uniwersum zmiennych stanu pokryto równomiernie gaussowskimi funkcjami przynależności o środkach w punktach $x \rightarrow \{-0.8;0;0.8\}$[m], $\theta \rightarrow \{-6;-1;0;1;6\}$[°], $\dot{x} \rightarrow \{-0.5;0;0.5\}$[m/s], $\dot{\theta} \rightarrow \{-50;0;50\}$[°/s].

Sygnał wzmocnienia przyjęto jako [1]:

$$r = \begin{cases} -1 & \text{gdy } |x| > 2.4m \text{ lub } |\theta| > 12^o \\ 0 & \text{w przeciwnym razie} \end{cases}$$

Założono, że każdy eksperyment uczenia pozwala na 100 krotny upadek wahadła, czemu każdorazowo będzie towarzyszyło wytworzenie sygnału wzmocnienia równego -1. Algorytm uczenia przerywano (przyjmując, że wagi są poprawnie nauczone),

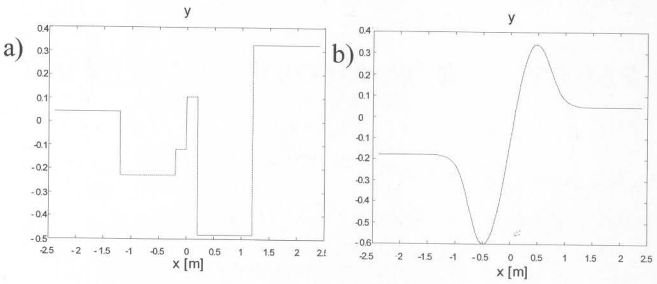
jeżeli wahadło nie upadło w 100000 kroków. Wszystkie zestawy nauczonych wag elementów ASE poddawano następnie testom w sterowaniu wahadłem odwróconym, podczas których nie prowadzono uczenia. Przyjęto, że jeżeli wahadło nie upadnie w czasie 2000 x 20ms = 40s, to taki zestaw wag zapewnia sukces w sterowaniu. Przykładowe przebiegi zmiennych stanu przedstawiono na rys. 3.



Rys. 3. Przebieg zmiennych stanu x i θ w trakcie sterowania za pomocą klasycznego elementu neuronowego ASE (a) i systemu T-S (b)
Fig. 3. The curve of the state variables x and θ for the control with classical neural element ASE (a) and T-S system

W przypadku sterowania za pomocą nauczonego elementu ASE (rozwiązanie klasyczne) otrzymane przebiegi mają charakter nietłumionych oscylacji (rys. 3a). Dynamika zmian zmiennej θ wynosi aż 0.352 radiana (20°). Na rysunku rys. 3b przedstawiono efekty sterowania przy użyciu nauczonego systemu rozmytego T-S, gdzie daje się zauważyć znaczne polepszenie charakteru zmian zmiennych stanu. I tak zmiana położenia wózka nabrała charakteru szybko tłumionych oscylacji - tak pożądanych w sterowaniu. Ponadto znacznie zmniejszyła się dynamika zmian kąta odchylenia wahadła od poziomu (tylko 0.0152 radiana, czyli mniej niż jeden stopień).

Przyczyną tych różnic wydaje się być zasada pracy poszczególnych systemów. W przypadku klasycznej sieci elementów ASE/ACE w danej chwili czasu aktywnym jest tylko jedno wyjście dekodera dyskretnego, co skutkuje wyznaczeniem sterowania na podstawie tylko jednego współczynnika wagowego elementu ASE. Zmiana wartości zmiennej stanu powodująca uaktywnienie innego wyjścia dekodera jest de-facto pobraniem innej wartości współczynnika wagowego (mogącego się znacznie różnić od poprzedniego). Różnica tych wartości jest powodem znacznej zmiany sterowania, co skutkuje „szarpnięciem" elementu wykonawczego, jakim jest wózek. Najlepiej charakter pracy tej sieci odda powierzchnia sterowania (rys. 4a).



Rys. 4. Powierzchnia sterowania $y = f(x, \dot{x} = 0, \theta = 0, \dot{\theta} = 0)$ dla klasycznej sieci ASE/ACE (a) i systemu rozmytego T-S (b)
Fig. 4. Control surface $y = f(x, \dot{x} = 0, \theta = 0, \dot{\theta} = 0)$ for the conventional ASE/ACE network (a) and fuzzy system T-S (b)

W przypadku systemu rozmytego specyfika pracy dekodera rozmytego powoduje, że wyznaczenie sterowania odbywa się na podstawie wielu aktywnych w danej chwili reguł sterowania. Aktywna jest nie tylko reguła o największym stopniu „odpalenia" (najbardziej predysponowana do opisu zaistniałej sytuacji), ale również reguły „sąsiednie". Ponadto w trakcie zmian wartości zmiennych stanu ma miejsce płynna zmiana stopnia aktywności reguł. W konsekwencji powierzchnia sterowania systemu rozmytego jest gładka (rys. 4b).

4. Podsumowanie

Zastosowanie algorytmu uczenia ze wzmocnieniem do systemu Takagi-Sugeno pozwoliło na pozyskanie reguł sterowania bez udziału eksperta, co jest niezwykle istotne w przypadku dużej ich liczby. Ponadto system rozmyty umożliwił uzyskanie oscylacyjnie tłumionych przebiegów położenia wózka, podczas gdy stosowanie klasycznej sieci neuronowej ASE/ACE skutkowało oscylacjami nietłumionymi. Dynamika zmian kąta odchylenia wahadła od pionu była aż 20-to krotnie mniejsza w przypadku systemu rozmytego. Reasumując, zastosowanie algorytmu uczenia ze wzmocnieniem do systemu Takagi-Sugeno wydaje się być jak najbardziej korzystnym z punktu widzenia zastosowań praktycznych.

5. Literatura

[1] A. G. Barto, R. S. Sutton, C. W. Anderson: Neuronlike adaptive elements that can solve difficult learning problem, IEEE Trans. SMC, 1983, 13, 834-847.

[2] H. R. Beom, H. S. Cho: A Sensor - Based Navigation for a Mobile Robot Using Fuzzy Logic and Reinforcement Learning, IEEE Trans. SMC, 1995, 25, 3, 464-477.

[3] C.-T. Lin: A neural fuzzy control system with structure and parameter learning, Fuzzy Sets Syst., 1995, 70, 183-212.

[4] R. Sutton, A. Barto: Reinforcement Learning: An Introduction, Cambridge, MA, MIT Press, 1998.

[5] C. Ye, N. Yung, D. Wang: A Fuzzy Controller With Supervised Learning Assisted Reinforcement Learning Algorithm for Obstacle Avoidance, IEEE Trans. SMC, 2003, 33, 1, 17-27.

[6] R. Zajdel: Algorytmy rozmyto-neuronowe i ich zastosowanie do sterowania małym robotem mobilnym, Rozprawa doktorska, Wrocław 1998.

[7] C. Zhou, Q. Meng: Dynamic balance of a biped robot using fuzzy reinforcement learning agents, Fuzzy Sets and Systems, 2003, 134, 169-187.

Title: Reinforcement learning with use of neuronlike elements ASE/ACE of Takagi-Sugeno controller

Artykuł recenzowany