

Keywords: Feedforward Net, Recurrent Net,
Deterministic Chaos

Tjaša Meško, Andrej Dobnikar
Fakulteta za elektrotehniko
in računalništvo, Ljubljana

V prispevku je analizirana sposobnost nevronske mreže pri napovedovanju determinističnega kaosa, ki je primer nelinearnega dinamičnega sistema. Opisana je primerjava med večnivojsko nevronske mreže s povratno povezavo in brez povratne povezave. Obema nevronske mreže med fazo učenja določimo število vozlov v skritem nivoju glede na hitrost konvergence. Izkazalo se je, da pri predikciji determinističnega kaosa navadna FF (feed forward) nevronska mreža ne zaostaja bistveno za rekurentno (recurrent), pri hitrosti učenja jo celo prekaša.

NEURAL NETWORKS FOR TIME SERIES PREDICTION. This contribution describes the results of testing a feedforward and a recurrent neural network on deterministic chaos prediction. Both networks are trained by the back-propagation algorithm which varies the number of hidden units depending on the convergence speed. It is shown that the feedforward network converges equally well as the recurrent one, even with a substantially lower number of hidden units.

1 Uvod

Pri obravnavanju naravnih pojavov ali pri iskanju kakšnih drugih zakonitosti se pogosto pojavi vprašanje napovedovanja prihodnosti. Kot primer bi lahko navedli vremensko napoved ali pa napovedovanje dogajanja na borzi. V primeru, da poznavanje sistema lahko zapišemo z rešljivo enačbo, lahko napovedujemo prihodnost, brž ko so določeni začetni pogoji. Če take enačbe ne poznamo, pa poiščemo empirične zakonitosti sistema in glede na le-te poskušamo bolj ali manj natančno napovedovati.

Pri slednji metodi se soočamo s problemi, kot je na primer šum. Periodičnost sistema je lahko zakrita s šumom v tolikšni meri, da se obnašanje sistema zdi popolnoma naključno.

Večnivojske nevronske mreže se v zadnjem času pogosto uporabljajo pri predikciji prihodnjih vrednosti časovnih serij s pomočjo učenja na preteklih primerih. Če se osredotočimo na časovne serije s šumom, je nevarnost, da se bo mreža primere in s tem tudi šum preveč naučila (*overfitting*). Pri tem je ključnega pomena velikost nevronske mreže (število vozlov v skritem nivoju).

2 Večnivojske nevronske mreže

Analiza in procesiranje časovnih serij sta mogoči na več načinov. V prispevku bomo obravnavali pristop z nevronskimi mrežami, kjer bomo uporabili dve znani topologiji: FF (*feed-forward*) in rekurentno (*recurrent*) topologijo. Naš cilj je primerjalno ovrednotiti obe topologiji predvsem z vidika hitrosti konvergence in velikosti topologije.

Obravnavali bomo le mreže z enim skritim nivojem. Mrežo učimo z back-propagation metodo [Rumelhart et al.,86]. Napaka mreže je definirana kot

$$E = \sum_n \sum_{i=1}^m (o_i^* - o_i)^2 \quad (1)$$

kjer je m število vozlov izhodnega nivoja, o_i je izhod i -tega izhodnega vozla, o_i^* je i -ta komponenta učilnega vektorja, n pa število elementov baze.

FF mreža z enim skritim nivojem je predstavljena na sliki 1. Izhod vozla i skritega (izhodnega) nivoja je definiran kot

$$o_i = f\left(\sum_j w_{ij}y_j + w_{i0}\right)$$

kjer je w_{ij} utež povezave med i -tim skritim (izhodnim) vozlom in j -tim vhodnim (skritim) vozlom, y_j je izhod vhodnega (skritega) vozla j in f je sigmoidna funkcija.

Primer rekurentne mreže oz. mreže s povratno povezavo dobimo, če FF mrežo z enim skritim nivojem dopolnimo s kontekstnim nivojem [Elman,90], ki je prav tako skriti nivo (glej sliko 2). Povezave od skritega do kontekstnega nivoja imajo konstantno utež 1, uteži ostalih povezav pa se med učenjem spreminjajo v skladu z back-propagation algoritmom. Izhode skritega nivoja si mreža torej v kontekstnem nivoju zapomni.

Pri učenju večnivojskih mrež se zmeraj pojavi problem optimalne konfiguracije mreže (število vozlov v skritem nivoju) pri dani velikosti dopustne napake (glej enačbo 1). V

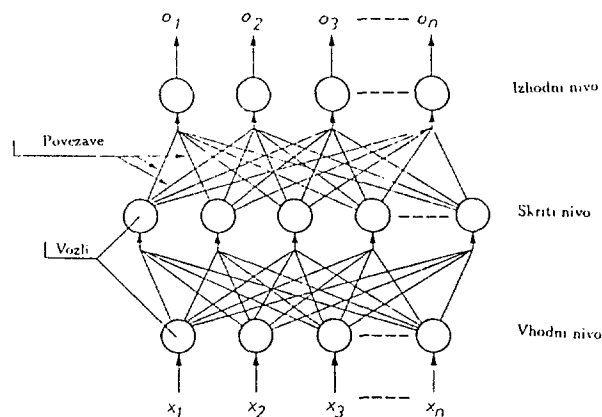


Figure 1: Mreža brez povratne povezave (FF).

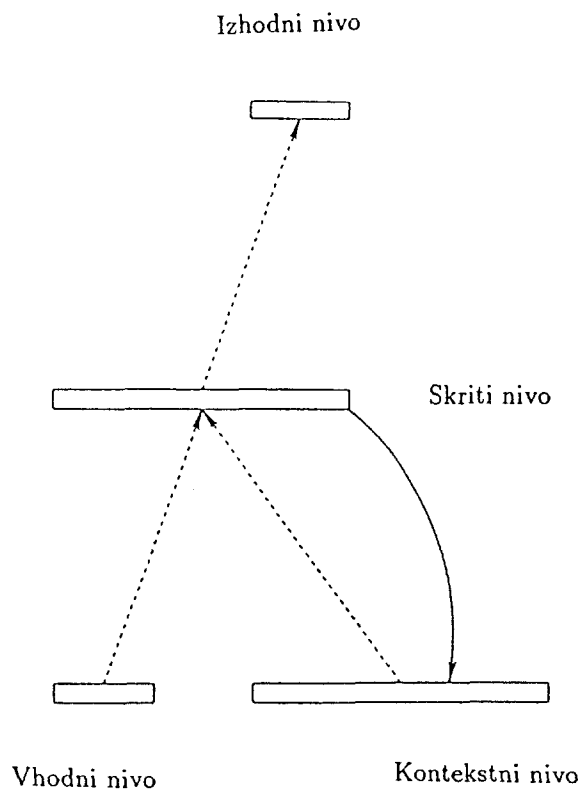


Figure 2: Mreža s povratno povezavo med skritim in kontekstnim nivojem (rekurentna mreža).

ta namen lahko pri učenju mrež z oz. brez povratne povezave uporabimo algoritem, ki spreminja število skritih vozlov [Hirose et al.,91] glede na hitrost padanja napake. Če napaka po nekem številu sprememb uteži ne pade za več kot 1%, dodamo vozle skritemu nivoju. Ko je napaka dovolj majhna, postopoma odvezemo vozle v skritem nivoju, dokler napaka spet ne naraste preko dovoljene meje.

3 Deterministični kaos

Za deterministični kaos sta pri predikciji časovnih serij značilni dve lastnosti [Weigend et al.,90]:

- Ker nemoč napovedovanja narašča eksponentno s časom, je možnost dolgoročne predikcije izključena.
- Mogoča je kratkoročna predikcija: časovne serije, ki se zdijo naključne, so lahko bile generirane z determinističnim sistemom.

Kaotično naravo predstavimo z vrednostjo x_t , ki je funkcija predhodnih d vrednosti časovne serije

$$(x_{t-1}, x_{t-2}, \dots, x_{t-d}) \mapsto x_t$$

Vektor $(x_{t-1}, x_{t-2}, \dots, x_{t-d})$ leži v d -dimenzionalnem prostoru časovne zakasnitve (*lag space*).

Standarden primer determinističnega kaosa je iterativna kvadratna preslikava na interval enote

$$x_t = 4x_{t-1}(1 - x_{t-1})$$

Ta preprosta enačba ilustrira nekatere ključne probleme determinističnega kaosa.

- Navidezna naključnost. Kljub temu, da je časovna serija generirana s pomočjo popolnoma determinističnega sistema brez šuma, se zdi sekvenca naključna.

- Kratkoročna predvidljivost. Ne glede na to, da se zdi časovna serija naključna, jo je mogoče dokaj natančno napovedati že s pomočjo prilegajoče se funkcije skozi množico zadnjih dveh točk $\{x_{t-1}, x_t\}$ iz preteklosti.
- Dolgoročna nepredvidljivost. V vsaki iteraciji izgubimo en bit informacije [Weigend et al.,90].

Deterministični sistemi se razlikujejo od stohastičnih po ciljih predvidevanja. Pri determinističnih sistemih želimo napovedati vrednost člena časovne serije v prihodnosti, ki naj bo čim bližje dejanski vrednosti v tistem trenutku. Pri stohastičnih sistemih pa lahko ima neko stanje več mogočih prehodov v različna stanja, ki se po vrednostih parametrov lahko precej razlikujejo. Predvidevanja so v tem primeru odvisna od verjetnosti stanj. Deterministični sistemi s šumom lahko predstavljajo poseben primer stohastičnih sistemov.

Če hočemo narediti model napovedovanja determinističnega sistema (lahko je kaotičen ali pa vsebuje šum), je potrebno izvesti naslednje tri točke:

1. Izberemo prostor časovne zakasnitve (*lag space*).
2. Aproximiramo predhodne časovne serije $\{x_t(x_{t-1}, x_{t-2}, \dots, x_{t-d})\}$ z gladko površino. Različni pristopi pri napovedovanju časovnih serij se razlikujejo predvsem po izbiri osnovnih elementov (polinomi, sigmoidi, ipd.).
3. Izberemo funkcijo cene, ki pove, kako dobra je aproksimacija.

V naslednjem poglavju bomo kot model napovedovanja kaotičnega sistema uporabili večnivojsko mrežo dveh topologij: FF in rekurentno.

4 Rezultati testov

Mreža s povratno povezavo in mreža brez povratne povezave sta imeli po en vhodni

	Učna baza	Testna baza
FF	0.0188	0.0192
Rekurentna	0.0186	0.0190

Table 1: *Napaka na učni in testni bazi pri uporabi večnivojske mreže z in brez povratne povezave pri kriteriju konvergence 0.02.*

	Učna baza	Testna baza
FF	0.0097	0.0098
Rekurentna	0.0098	0.0520

Table 2: *Napaka na učni in testni bazi pri uporabi večnivojske mreže z in brez povratne povezave pri kriteriju konvergence 0.01.*

in en izhodni voz. Število vozlov skritega nivoja smo določili z algoritmom, ki spreminja število skritih vozlov [Hirose et al.,91] glede na hitrost padanja napake. Pri rekurentni mreži je število skritega nivoja in konteksta zmeraj enako, zato ju povečamo ali zmanjšamo naenkrat. Pri testu obeh mrež je bil uporabljen deterministični kaos z enačbo

$$x_t = 3.8x_{t-1}(1 - x_{t-1})$$

pri čemer je $x_0 = 0.5$. Deterministični kaos smo izbrali zato, ker ga je relativno preprosto generirati.

Učna baza je bila sestavljena iz prvih stotih elementov časovne serije, testna baza pa iz nadaljnjih stotih elementov. Vsota kvadratov napak (glej enačbo 1), za vsak primer posebej na testni in učni bazi, je podana v tabeli 1 in 2. Razvidno je, da je razlika med napako na testni in na učni bazi relativno majhna, kar je posledica dejstva, da ni bilo šuma. Opazna razlika med napako na testni in učni bazi je le pri rekurentni mreži pri velikem številu iteracij.

Pri prvem poskusu je bila vrednost vsote kvadratov napak 0.02 pogoj za konvergenco na učni bazi. Primerjavo med nevronskima mrežama vidimo na sliki 3. Pri drugem poskusu je bil parameter konvergence

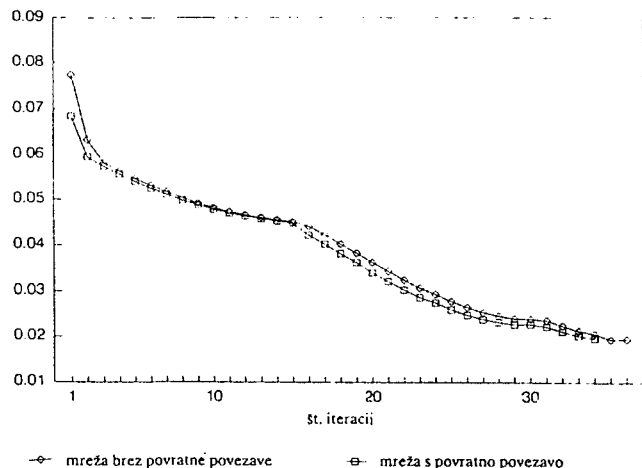


Figure 3: *Primerjava napake pri večnivojski mreži z in brez povratne povezave. Parameter konvergence ima vrednost 0.02.*

zmanjšan na 0.01. To je povzročilo precej daljše učenje in tudi povečalo število vozlov v skitem nivoju (glej sliko 4). Za dosego kriterija konvergence 0.02 sta pri obeh topologijah zadoščala dva vozla v skitem nivoju (pri rekurentni še dva vozla v kontekstu). FF mreža je zmanjšala napako pod 0.02 po 36 iteracijah, rekurentna mreža pa po 34 iteracijah. Za dosego napake, manjše od 0.01, je FF potrebovala 21 skritih vozlov in 84 iteracij, rekurentna pa kar 23 vozlov v skitem nivoju in 23 vozlov v kontekstu ter 87 iteracij.

5 Zaključek

Mreža brez povratne povezave kaže na primeru determinističnega kaosa glede na obseg topologije in konvergenco ugodnejše lastnosti kot rekurentna mreža. Pri majhnem številu iteracij je napaka rekurentne mreže nekaj manjša od napake FF, vendar se ta prednost s številom iteracij izgubi. Učenje rekurentne mreže je tudi precej bolj zamudno zaradi kontekstnega nivoja. FF mreža dobro posplošuje z učne na testno bazo ne glede na število iteracij, rekurentna pa posplošuje pri

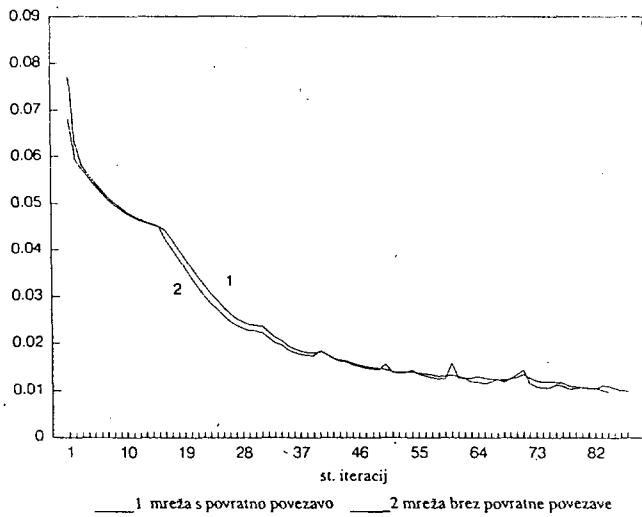


Figure 4: *Primerjava napake pri večnivojski mreži z in brez povratne povezave. Parameter konvergence ima vrednost 0.01.*

večjem številu iteracij slabše.

Literatura

- [1] Elman, J.L. (1990). "Finding Structure in Time", *Cognitive Science* 14, 179-211, 1990.
- [2] Hirose, Y., Yamashita, K. and Hijiya, S. (1991). "Back-Propagation Algorithm Which Varies the Number of Hidden Units", *Neural Networks*, Vol.4, pp.61-66, 1991.
- [3] Rumelhart, D., Hinton, G. and Williams, R. (1986). "Parallel distributed processing: exploration in the microstructure of cognition", Vol. 1, MIT Press.
- [4] Weigend, A.S., Huberman, B.A. and Rumelhart, D.E. (1990). "Predicting the Future: A Connectionist Approach", *International Journal of Neural Systems*, Vol.1, No.3, pp.193-209, 1990.