

NO HOPFĪLDA TĪKLA LĪDZ BOLCMANA MAŠĪNAI

Pēteris Grabusts
Rēzeknes Augstskola
90 Atbrīvošanas aleja, Rēzekne LV-4600, Latvija
Tālrunis: (46)23798
Fax: (46)23709
E-mail: peter@cs.ru.lv

Anotācija

Sekojošais materiāls dod ieskatu mākslīgo neironu tīklu klasifikācijā un apskata vienu no pazīstamākajām neironu tīkla arhitektūrām - Hopfilda tīklu un tā modifikāciju - Bolcmana mašīnu.

1. Ievads

Ar mākslīgo neironu tīklu saprot informācijas apstrādes sistēmu, kas sastāv no liela skaita paralēli strādājošu elementu - neironu, kuri saistīti savā starpā ar informācijas pārraides kanāliem - saitēm jeb sinapsēm

2. Mākslīgo neironu tīklu klasifikācija

2.1. Pamatjēdzieni

Mākslīgo neironu tīklu pamatelements ir neirons, ko bieži, atšķirībā no bioloģiskā neirona, sauc par formālo vai tehnisko neironu.

Formālajam neironam (skat. Zīm. 1.) ir sekojoša struktūra :

- 1) Neironam ir n ieejas kanāli un viens izejas kanāls;
- 2) Katram ieejas kanālam x_i atbilstībā tiek piešķirts skaitlis w_i , ko sauc par šī kanāla svaru. Svaru vektors ir $\mathbf{w}=\{w_1, w_2, \dots, w_n\}$.
- 3) Visu ieeju svērtie signāli tiek summēti :

$$a = \sum_i w_i x_i \quad (1)$$

4) Summārais signāls tiek modificēts ar pārvades jeb aktivācijas funkciju un padots uz izejas kanālu.

Kā pārvades funkcijas var tikt izmantotas :

a) *lineārā* $y(a)=a$ (2)

b) *sigmoidālā(logistiskā)* $y(a)=\frac{1}{1+e^{-a}}$ $y \in (0,1)$ (3)

c) *sigmoidālā(tanh)* $y(a)=\tanh(a)$ $y \in (-1,1)$ (4)

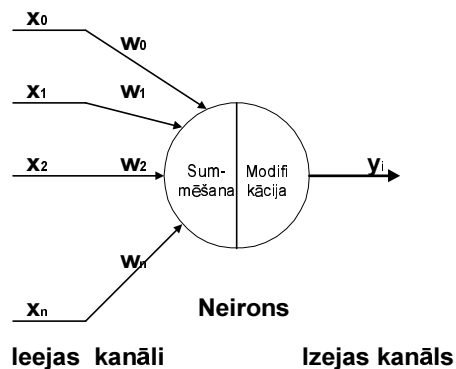
d) *sliedzīga* $y(a)=\theta(a)=\begin{cases} +1 & a > 0 \\ -1 & a \leq 0 \end{cases}$ (5)

$$e) \text{ stohastiskā} \quad y(a) = \begin{cases} +1 \\ -1 \end{cases} \quad \text{ar } p = \frac{1}{1 + e^{-a}} \quad \text{cit'adi} \quad (6)$$

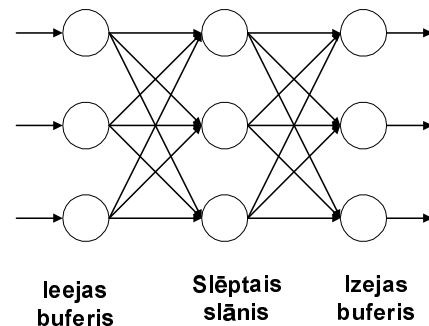
f) u.c.

Neironu tīklam (skat. piemēru. Zīm. 2.) ir sekojoša struktūra :

- 1) tīkls sastāv no daudziem savā starpā saistītiem neironiem;
- 2) neironi sakārtoti līmeņos jeb slāņos;
- 3) pirmais slānis ir ieejas buferis, kas ievada signālu vektoru tīklā;
pēdējais slānis ir izejas buferis, kas nosaka tīkla reakciju uz ieejas signāla vektoru;
starp pirmo un pēdējo slāni var būt viens vai vairāki slēptie slāņi (starp-slāņi), kuri veic tīkla darbības rezultātu uzkrājēja funkcijas;
- 4) starpslāņu savienojumi var būt pilni, neejauši vai pa elementu grupām;
- 5) tīklu ar slāņiem un signālu apstrādes funkcijām sauc par perceptronu.



Zīm. 1.



Zīm. 2.

Neironu tīkla funkcionēšanā būtiska nozīme ir tīkla apmācībai t.i. svaru adaptācijas procesam, veidojot neironu reakcijas uz ieejas signālu vektoru līdz pat izejas buferim. Izejas bufera signāliem būtu jāsakrīt ar vēlamo rezultātu. Tīkla funkcionēšana uzlabojas piemērojot svaru koeficientus.

Apmācības procesā būtu jāzina:

- 1) **tīkla funkcionēšanas modelis** - t.i. jāzina tīklam pieejamā informācija. To nosaka tīkla apmācīšanas veids.

Izšķir :

- a) *apmācība ar skolotāju*-izejas vektoram jāsakrīt ar ieejas signālu vektoru;
- b) *apmācība bez skolotāja*(pašapmācīšanās) - vēlamās izejas nav zināmas;
- c) *jauktā apmācība*-svāri tiek modificēti gan ar skolotāja apmācību, gan bez.

- 2) **apmācības likums** - t.i. tīkla svaru modifikācijas algoritms.

Izšķir 3 apmācības likumu tipus :

- a) *kļūdu korekcijas (delta) likums*.

Apmācībā ar skolotāju katram ieejas piemēram uzdod vēlamu izeju **d**. Reālā tīkla izeja **y** var nesakrist ar vēlamo. Signālu **d-y** izmanto svaru modifikācijai, pakāpeniski samazinot kļūdu.

- b) *Heba apmācība*.

Ieejas svars tiek palielināts, ja ieejas signāla līmenis ir augsts un arī vēlamā izejas signāla līmenis ir augsts.

c) *Apmācība konkurences apstākļos.*

Neironi sacenšas savā starpā par tiesībām aktivizēties. Šāds paņēmieni zināms ar nosaukumu “uzvarētājs ņem visu”.

Dažādos literatūras avotos uzrādīts vēl viens apmācības likums :

d) *Boicmana apmācība.*

Svaru koeficienti tiek adaptēti tādā veidā, lai redzamo neironu stāvoklis atbilstu vēlamajam varbūtību sadalījumam.

2.2. Neironu tīklu arhitektūra

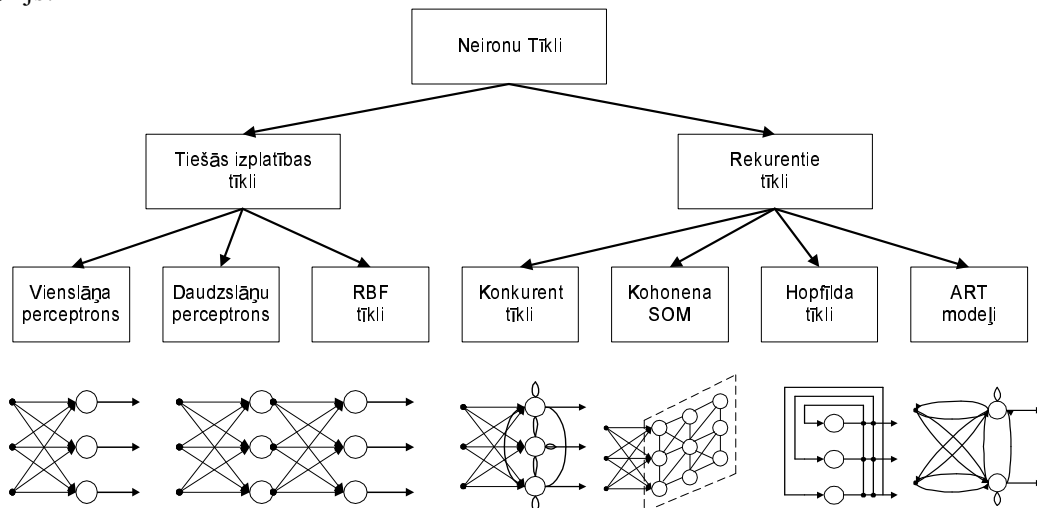
Mākslīgo neironu tīklu var uzdot kā grafu ar svērtām saitēm, kuros neironi ir mezgli. Pēc saišu veida neironu tīkli var tikt iedalīti divās klasēs (skat. zīm. 3.):

1) *tiešo saišu tīkli*, kuros grafiem nav cilpu;

Tie ir vienvirziena tīkli, kuros informācijas apstrāde notiek no ieejas bufera caur starpslāņiem līdz izejas buferim vienā virzienā.

2) *rekurentie tīkli jeb tīkli ar atgriezeniskām saitēm.*

Informācija katrā slānī vai starp slāņiem cirkulē tik ilgi, kamēr izpildās konverģences kritērijs.



Zīm. 3. Neironu tīklu klasifikācija.

Vienslāņa perceptrons - spēj apmācīties klasificēt ieejas tēlus, kuri ir lineāri atdalāmi. Apmācībā izmanto perceptrona apmācības algoritmu.

Daudzslāņu perceptrons - viens no populārākajiem tīkliem lietotāju vidū. Katrs neirons aktivizējas ar sigmoidālās vai sliekšņa funkcijas palīdzību. Apmācībā var izmantot kļūdas atgriezeniskās izplatības algoritmu, Adaline vai Madaline algoritmus. Veicamie tīkla uzdevumi - tēlu klasifikācija, funkciju aproksimācija, prognozēšana, vadība.

RBF tīkli - tīkls ar radiālām aktivizācijas funkcijām. Katrs slēptā slāņa neirons aktivizējas ar radiālās gausa tipa funkcijas palīdzību. Radiālā aktivizācijas funkcija (kodola funkcija) centrējas punktā, kuru nosaka neironu svaru vektors. Kodola funkcijas

pozīcijai jābūt apmācītai ar piemēriem. Katrs izejas elements izskaitļo šo RBF lineāro kombināciju. Apmācībā izmanto dažādus RBF tīkla apmācības algoritmus. Pamatalgoritms izmanto jaukto apmācību.

Kompetitīvie jeb konkurences tipa tīkli - Izejas bufera neironi sacenšas par tiesībām reaģēt uz ieejas vektoru. Uzvar tas neirons, kura reakcija ir visspēcīgākā, tas tiek modificēts, lai iegūtu lielāku līdzību ar doto ieejas vektoru. Apmācība dod iespēju klasterizēt ieejas datus : līdzīgie piemēri grupējas un tiek uzdoti kā viens piemērs. Tīkls tiek izmantots datu analīzes uzdevumos.

Kohonena SOM jeb Kohonena kartes - Konkurences tipa tīklu speciāls gadījums, kurā tiek definēts “telpisks” apgabals katram izejas elementam. Apgabals var būt kvadrāts, taisnstūris vai riņķis. Sākotnēji apgabala izmērs ir robežās no 1/2 līdz 2/3 no tīkla izmēra un samazinās pēc noteikta likuma. Apmācības laikā tiek modificēti visi svāri, kas saistīti ar “uzvarētāju”. Izmanto datu analīzē.

Hopfilda tīkls - tīks apskatīts nākošajā punktā.

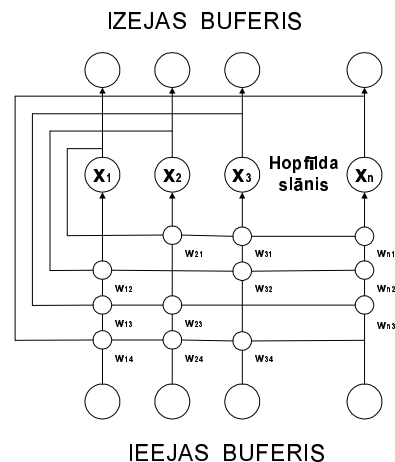
ART modeļi jeb adaptīvās rezonanses modeļi - Apskata attēlu klasifikācijas bez ārējā skolotāja pašorganizācijas jautājumus. Apmācības algoritms koriģē esošo kategorijas prototipu, ja ieejas vektors tam ir atbilstošs. Šajā gadījumā tie nonāk rezonanses stāvoklī. Sakrišanas pakāpe raksturojas ar līdzības koeficientu, kas saistīts arī ar kategorijas skaitli. Ja ieejas vektors nav pietiekoši līdzīgs nevienam esošam tīkla prototipam, veidojas jauna kategorija. Šāds tīkls veic datu analīzes uzdevumus.

3. Hopfilda tīkls kā asociatīvā atmiņa

Visu neironu asociatīvos tīklus var iedalīt divās grupās : **heteroasociatīvie** un **autoasociatīvie** tīkli. Heteroasociatīvajos tīklos ieejas vektors **X** asociējas ar izejas vektoru **Y** t.i. apmācītam tīklam ieejā padodot vektoru **X** - izejā tiek izdots vektors **Y**. Autoasociatīvie tīkli asociē ieejas vektoru **X** pašu ar sevi t.i. padodot ieejā vektora **X** daļu , atjauno to. Autoasociatīvie tīkli parasti ir vienslāņa, ar atgriezeniskām saitēm starp neironiem. Faktiski autoasociatīvajos tīklos ieejas slānis vienlaikus ir arī izejas slānis.

Tā kā iepriekš nevar paredzēt, kādas asociācijas var rasties starp objektu pazīmēm, tad tīklam jābūt savstarpēji pilnībā saistītam t.i. katrs neirons ir saistīts ar katru tīkla neironu (izņemot pašam ar sevi). Šādu tīklu nosauca par **Hopfilda tīklu**(skat. Zīm.4). Tas ir viens no autoasociatīvo neironu tīklu realizācijas veidiem.

Saites starp neironiem ir simetriskas t.i. $w_{ij}=w_{ji}$ un $w_{ii}=0$.



Zīm. 4. Hopfilda tīkla arhitektūra.

Hopfīlda tīklam ir divas iezīmes, kas atšķir to no citiem asociatīvajiem neironu tīkliem:

- 1) katrā laika momentā savu stāvokli var izmainīt tikai viens elements (asinhrona stāvokļa maiņa);
- 2) katrs elements var saņemt signālu ne tikai no citiem tīkla neironiem, bet arī no papildus vadības ieejas, kas atļauj vai aizliedz aktivizāciju.

Asinhronā elementu izmaiņa ļauj ievest funkciju, ko Hopfīlds nosauca par **enerģijas** funkciju. Galvenā enerģijas funkcijas īpašība ir tā, ka tā samazinās un sasniedz lokālo minimumu (**atraktoru**), kur tīklam ir pastāvīga enerģija.

Ja tīklā glabājamie tēlu vektori ir atraktori, šādu tīklu var uzskatīt par asociatīvo atmiņu. Asociatīvajai atmiņai parasti izšķir divus režīmus: **apmācība** un **atcerēšanās**. Apmācības saišu svāri tiek noteikti tā, lai atraktori iegaumētu saglabājamus vektorus $\{x_1, x_2, \dots, x_n\}$. Atcerēšanās režīmā ieejas piemērs tiek izmantots kā tīkla sākuma stāvoklis. Izejas vektors tiek izdots tad, kad tīkls sasniedz līdzsvaru.

Apmācības likums.

Lai tīkls varētu strādāt, tajā jānosaka saites (asociācijas) starp objektiem. Tiek pielietota vai nu apmācība bez skolotāja (**Heba** apmācības likums) vai apmācība ar skolotāju (**Vidrova-Hofa** apmācības likums).

Ieejas signālu aprakstām ar vektoru $\mathbf{X}(p)$, kur $p=1, 2, \dots, k$ un

$$\mathbf{X}(p) = \{X_1(p), \dots, X_i(p), \dots, X_n(p)\} \quad (7)$$

Katra elementa x_i vērtība var būt $+1$ vai -1 . Tad Heba apmācības likums diskrētām Hopfīlda tīklam svarus nosaka sekojošā veidā:

$$W_{ij} = \sum_{p=1}^k (2x_i - 1)(2x_j - 1) \quad i \neq j \quad \text{un} \quad w_{ij} = w_{ji}, \quad w_{ii} = 0 \quad (8)$$

Hopfīlda tīkla darbības algoritms (jeb praktiski to var nosaukt par **atcerēšanos**) ir sekojošs:

1. Uz ieeju padod nezināmu signālu y_i , t.i. faktiski notiek tīkla aktivizācija: $x_i = y_i$
2. Izskaitļo jauno neironu stāvokli (p - iterācijas numurs):

$$S_i(p+1) = \sum_{j=1}^n w_{ij} y_j(p) \quad (9)$$

$$\text{un izejas signālu} \quad X_i(p+1) = f[S_i(p+1)], \quad (10)$$

kur f ir kāda no pārvades funkcijām (2)-(6).

3. Pārbauda, vai izejas vērtības ir pamainījušās. Ja pamainījās - iet uz p.2., pretējā gadījumā ir algoritma darba beigas t.i. tīkls ir apmācījies.

Atcerēšanās procesā stāvokļi tiek modificēti tā, lai enerģija samazinātos tāpēc tajā tīkla daļā, kurā kā atmiņā nostiprinās kādi tēli, parādās lokāli enerģijas minimumi. Neironu tīkla atmiņas asociativitāte izpaužas tā, ka uzrādot apmācāmajam tīklam nepilnīgu ieejas vektoru, tīkls atmiņā atradīs tam tuvāko pilno vektoru un izdos tam atbilstošu izeju.

Hopfīlds parādīja, ka tīkls konverģēs uz stāvokli, kad visu tīkla neironu izejas stāvokļi neizmainīsies. Tīklam pēc apmācības ir stabili neironu stāvokļi, kuros enerģijas funkcijai ir lokāli minimumi.

$$E = -\frac{1}{2} \sum_{i < j} \sum w_{ij} x_i x_j \quad (11)$$

Mainoties viena apstrādājamā elementa stāvoklim, mainīsies arī tīkla enerģija:

$$\Delta E_i = -\frac{1}{2} \Delta x_i * \sum_j w_{ij} x_j \quad (12)$$

Jebkura stāvokļa maiņa samazinās tīkla enerģijas līmeni t.i. tīkls garantēti konverģēs pēc galīga skaita iterāciju.

4. Bolcmana mašīna

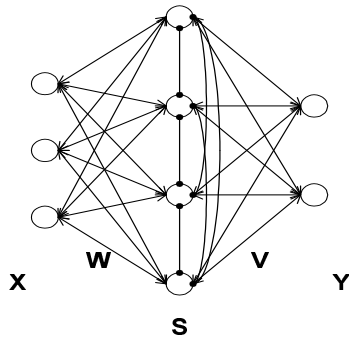
Hopfīlda tīklos tika ieviesta enerģijas funkcija un tika parādīts, ka tīkls var nonākt lokālā enerģijas minimuma stāvoklī t.i. tīkls konverģē uz kādu lokālu minimumu, bet neatrod globālu minimumu. Vienkāršākais veids, kā izkļūt no lokālā minimuma ir “pārlekt” stāvoklī ar augstāku enerģiju.

Kā viens no Hopfīlda tīklu turpinājumiem ir Bolcmana modelis jeb Bolcmana mašīna (skat. zīm. 5). Bolcmana mašīnu sāka pētīt Aklejs, Hintons un Sejnovskis 1985. gadā. Bolcmana mašīnā tiek vilktas paralēles ar pakāpenisku “termālu atdzišanu”.

Līdzīgi Hopfīlda tīklam, Bolcmana tīkls meklē stāvokli ar zemāko enerģiju. Kad svērto svaru summa aktīvajam neironam ir pozitīva, tad neirons ir aktīvs. Neirons aktivizējas ar sekojošu varbūtības :

$$p = \frac{1}{1 + e^{-(\Delta E)/T}} \quad (13)$$

kur T - “mākslīga temperatūra” un E - aktīvo savienojumu summa.



Zīm. 5. Bolcmana mašīnas piemērs.

Bolcmana apmācība tiek veikta vairākos etapos :

1. “Mākslīgai temperatūrai” T piešķir lielu sākumvērtību.
2. Caur tīklu izlaiž ieejas vektorus un izskaitļo izejas vērtības un mērķa funkciju.
3. Pamaina svaru vērtības un pārrēķina tīkla izejas vērtības un mērķa funkciju.
4. Ja mērķa funkcijas vērtība samazinājās (t.i. uzlabojās), saglabā svaru izmaiņas.

Ja svaru izmaiņas noved pie mērķa funkcijas palielināšanās, tad šo izmaiņu saglabāšanas varbūtību izskaitļo ar Bolcmana sadalījuma palīdzību :

$$p(c) = e^{-c/kT} \quad (14)$$

kur p - mērķa funkcijas izmaiņas varbūtība, k - konstante (analogā Bolcmana konstantei), T - mākslīga temperatūra.

Izvēlas gadījuma skaitli $r \in [0,1]$. Ja $p > r$, tad izmaiņas saglabājas. Pretējā gadījumā atstāj iepriekšējo svaru vērtību. Šī procedūra ļauj sistēmai veikt varbūtisku soli virzienā, kas dod iespēju tai izrauties no lokālajiem minimumiem. Soļi 3 un 4 tiek atkārtoti katram no tīkla svāriem, pakāpeniski samazinot temperatūru T , kad būs sasniegta pietiekami zema mērķa funkcijas vērtība. Tad tiek piedāvāts cits ieejas vektors un apmācības process turpinās. Tīkls apmācās no visiem ieejas vektoriem, kamēr mērķa funkcija netiks pielāgota katram no tiem.

Bolcmana mašīnu izmanto tēlu atpazīšanas un klasifikācijas uzdevumos. Trūkums - lēns apmācības algoritms, taču tas ļauj izkļūt no lokālajiem minimumiem.

Secinājumi

Protams, pastāv arī citas mākslīgo neironu tīklu klasifikācijas metodes, kā arī vesela virkne apmācības algoritmu, kas šajā rakstā netiek aplūkoti. Var minēt BSB tīklus, neironu tīklus ar informācijas kontrizplatību (CPN tīkli), divvirzienu asociatīvā atmiņa (BAM) un virkni citu.

Ieteicamā literatūra

Mākslīgie neironu tīkli: arhitektūra, algoritmi un pielietojumi.(1998).Mācību līdzeklis. Rīga.

Alexander, I., Morton, H. (1991) An Introduction to Neural Computing. Chapman & Hall, London.

Fausett, L. (1994) Fundamentals of Neural Networks: Architectures, Algorithms and Applications. Prentice Hall International Inc.

Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl.Acad.Sci. USA*, **79**, 2554-2558.

Ackley, D.H., Hinton, G.E. and Sejnowski, T.J. (1985) A learning algorithm for Boltzmann machines. *Cognitive Science*, **9**, 147-169.