

Vizualizavimui skirto neuroninio tinklo mokymosi greičio optimizavimas

Viktor MEDVEDEV, Gintautas DZEMYDA (MII)

el. paštas: viktor.m@ktl.mii.lt, dzemyda@ktl.mii.lt

1. Įvadas

Požymių išskyrimas ir daugiamačių duomenų projektavimas yra svarbios problemos vektorių atpažinimo ir tyrimo analizėje. Straipsnyje nagrinėjamas dirbtinių neuroninių tinklų panaudojimas daugiamačių duomenų vizualizavimui.

Nustatant ryšius tarp klasikinių metodų ir neuroninių tinklų, buvo pastebėta, kad kai kurie neuroniniai tinklai faktiškai realizuoja požymių išskyrimo ir duomenų projektavimo algoritmus. Atliekant tyrimus, keletas klasikinių požymių išskyrimo ir duomenų projektavimo metodų buvo realizuoti naudojant neuroninių tinklų struktūras [1, 6]. Tokie tinklai dažnai yra pranašesni už tradicinius metodus. Pavyzdžiui, SAMANN neuroninis tinklas, sukurtas Sammono netiesinės projekcijos algoritmui, suteikia apibendrinimo galimybę naujų duomenų vizualizavimui. O tai yra savybė, kurios nėra originaliajame algoritme. Šio darbo tikslas – nustatyti optimalią mokymosi parametro reikšmę, kad mokymo algoritmas greičiau konverguotų.

2. Daugiamačių duomenų vizualizavimas

2.1. Sammono projekcija

Vienas iš daugiamačių duomenų vizualizavimo metodų yra Sammono algoritmas [4] Sammono projekcija yra netiesinio daugelio kintamųjų vektorių atvaizdavimo į mažesnio matavimo erdvę metodas.

Tarkime, kad turime N vektorių n -matėje erdvėje $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, $i = 1, \dots, N$ ir atitinkamai apibrėšime N vektorių d -matėje erdvėje ($d < n$, paprastai $d = 2$ arba 3) $Y_i = (y_{i1}, y_{i2}, \dots, y_{id})$, $i = 1, \dots, N$. Atstumą tarp vektorių X_i ir X_j n -matėje erdvėje pažymėsime d_{ij}^* , ir atstumą tarp atitinkamų vektorių Y_i ir Y_j d -matėje erdvėje pažymėsime d_{ij} [4, 7]. Sammono klaida (1) – tai matas, kuris parodo, kaip tiksliai atstumai tarp vektorių išlaikomi pereinant iš didesnio matavimo erdvės į mažesnio matavimo erdvę:

$$E = \frac{1}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n d_{ij}^*} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*}. \quad (1)$$

2.2. SAMANN algoritmas

Sammono algoritmo trūkumas yra tas, kad jis netinka tuo atveju, kai reikia greitai ir tiksliai atvaizduoti naujus vektorius. Atėjus naujam vektoriui, visa mokymo procedūra turi būti pakartota iš naujo, t.y. neuroninio tinklo mokymui naudojama sena duomenų aibė ir nauji duomenys. Mao ir Jain (1995) šios problemos sprendimui pasiūlė SAMANN tinklą [6]. SAMANN – tai specifinė neuroninio tinklo mokymo taisyklė (klaidos sklidimo atgal būdu), leidžianti įprastam tiesioginio sklidimo daugiasluoksniams dirbtiniam neuroniniam tinklui realizuoti Sammono projekciją mokymo be mokytojo būdu. SAMANN algoritmas turi apibendrinimo galimybę dirbant su naujais vektoriais. 1 pav. pateiktas tokio tinklo modelis.

Tegul $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, $i = 1, \dots, N$ yra n -mačių įėjimo vektorių aibė. Šie vektoriai naudojami tinklui mokyti, siekiant gauti jų d -matę projekciją ($d < n$). Vizualizavimo erdvės dimensija yra d , $d < n$. Pažymėkime j -otojo elemento l -tajame sluoksnyje išėjimą $y_j^{(l)}$, $j = 1, 2, \dots, n_l$, $l = 0, 1, 2, \dots, L$; čia n_l yra sluoksnio l paslėptų neuronų skaičius, L – sluoksnių skaičius, ir $y_j^{(0)} = X_j$, $j = 1, 2, \dots, d$. Svorį jungtyje tarp i -tojo neurono $(l - 1)$ -ajame sluoksnyje ir j -tojo neurono l -tajame sluoksnyje žymėsime $\omega_{ij}^{(l)}$. $\omega_{0j}^{(l)}$ – j -tojo neurono l -tajame sluoksnyje slenksčio reikšmė, $y_0^{(l)} = 1.0$. Kiekvieno neurono išėjimo reikšmei skaičiuoti naudosime sigmoidinę funkciją $g(h)$, kurios išeinamųjų reikšmių intervalas $(0, 1)$; čia h – visų elementų svorių suma. j -tojo elemento l -tajame sluoksnyje išėjimas užrašomas taip:

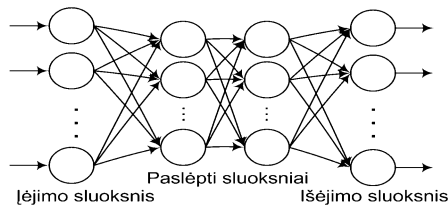
$$y_j^{(l)} = g\left(\sum_{i=0}^{n_l} \omega_{ij}^{(l)} y_i^{(l-1)}\right), \quad l = 1, 2, \dots, L. \quad (2)$$

Bendros svorių atnaujinimo taisyklės kiekvienam paslėptam sluoksniui ($l = 1, \dots, L - 1$) ir išėjimo sluoksniui ($l = L$):

$$\Delta\omega_{ij}^{(l)} = -\eta \frac{\partial E_{\mu\nu}}{\partial \omega_{ij}^{(l)}} = -\eta \left(\Delta_{ij}^{(l)}(\mu) y_i^{l-1}(\mu) - \Delta_{ij}^{(l)}(\nu) y_i^{l-1}(\nu) \right), \quad (3)$$

η – mokymosi parametro reikšmė.

Kiekviename mokymo žingsnyje į neuroninį tinklą paduodami du taškai. Skaičiuojamas atstumas tarp neuroninio tinklo vektorių išėjimų ir nustatomas klaidos matas,



1 pav. Tiesioginio sklidimo 3-jų sluoksnių neuroninis tinklas Sammono projekcijai.

naudojant šį atstumą ir atstumą tarp taškų įėjimo erdvėje. Šios klaidos pagrindu gaunama svorių atnaujinimo taisyklė. Norint atnaujinti svorius ω , į neuroninį tinklą tuo pat metu turim paduoti vektorių porą μ ir ν , vietoj vieno vektoriaus, kaip sklidimo atgal mokymo algoritme. Norint tai padaryti, galima arba sukonstruoti du identiškus tinklus, arba tiesiog laikyti atmintyje visus pirmojo vektoriaus išėjimo reikšmes, prieš paduodant antrąjį vektorių. Neuroninis tinklas apmokamas jam daug kartų pateikiant skirtingas apmokymo aibės vektorių poras. Mokymo metu perskaičiuojami neuroninio tinklo svoriai. Kadangi šiam algoritmui nereikalingi apmokymo vektoriai, tai šis algoritmas yra nekontroliuojamas.

SAMANN algoritmas atrodo taip: 1) atsitiktinai nustatomi SAMANN tinklo svoriai; 2) atsitiktinai parenkami du vektoriai ir paduodami kiekvienam iš tinklų po vieną (pirmam tinklui – pirmas vektorius, antram – antras); 3) atnaujinami svoriai, sklidimo atgal būdu, pradedant nuo išėjimo sluoksnio; 4) kartojami žingsniai 2–3 kol bus pasirinktos ir tinklams paduotos visos galimos vektorių poros; 5) skaičiuojama Sammono klaida; 6) jeigu Sammono klaidos reikšmė mažesnė už pasirinktą slenkstį arba iteracijų skaičius M (2–5 žingsniai) viršija maksimalų leistiną skaičių, tuomet sustojam; priešingu atveju vėl pradedam nuo 2 žingsnio.

3. Mokymosi parametro optimizavimas

Eksperimentams buvo naudojamos trys duomenų aibės:

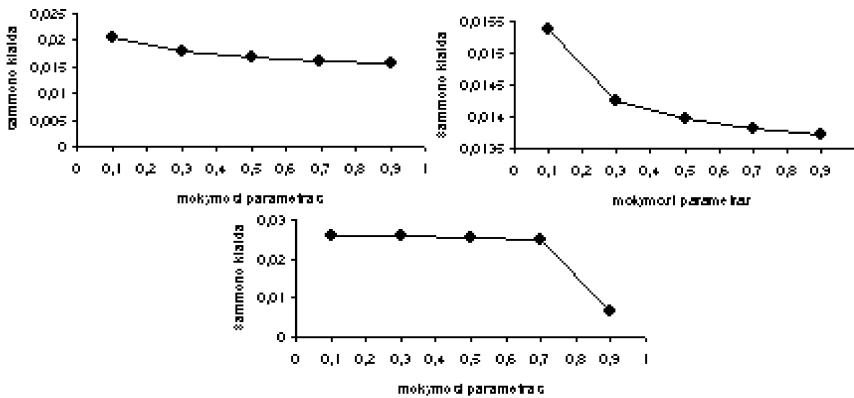
- Iris duomenų aibė [8]. Aibė (dar žinoma kaip Fišerio Iris aibė) sudaro 150 keturmačių vektorių (3 klasės po 50 vektorių);
- Salinity duomenų aibė [10]. Tai vandens sūrumo (druskos koncentracija) matavimai. Aibė sudaro 28 keturmačiai vektoriai;
- HBK duomenų aibė [9] (dirbtinė duomenų aibė). Ją sudaro 75 keturmačiai vektoriai.

Tinklo mokymas turi būti efektyvus ir mokymo algoritmas turi greitai konverguoti. Nustatyta, jog SAMANN neuroninio tinklo mokymui reikia daug skaičiuojamųjų sąnaudų, todėl naujus svorius ir tikslią duomenų projekciją siekiama gauti per trumpą laiką. Analizuojant SAMANN tinklą pastebėta, kad projekcijos paklaida priklauso nuo skirtingų parametrų. Paskutiniai tyrimai parodė, kad norint pasiekti gerų vizualizavimo rezultatų, reikia teisingai parinkti mokymosi parametro η reikšmę. Iki šiol buvo teigiama [6], kad vizualizavimas duoda geriausių rezultatų, kai reikšmė η imama iš intervalo (0, 1). J. Mao ir K. Jain [6] savo tyrimuose naudojo $\eta = 0, 7$. Tačiau tuo atveju tinklo mokymas vyksta labai lėtai. Viena iš galimų to priežasčių yra tai, kad SAMANN tinklo atveju intervalas (0, 1) nėra pats geriausias. Taigi tikslinga ieškoti optimalios mokymosi parametro reikšmės, kuri nebūtina turi būti intervale (0, 1).

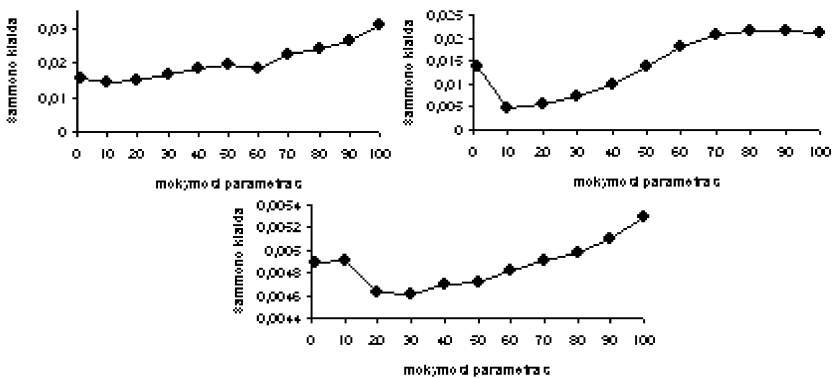
Analizuojant SAMANN algoritmą, buvo nagrinėjamas vienasluoksnis tiesioginio sklidimo dirbtinis neuroninis tinklas, turintis vieną paslėptą sluoksnį ir du išėjimus ($d = 2$). Visais atvejais buvo imamas vienodas paslėptojo sluoksnio neuronų skaičius ($n_2 = 20$), iteracijų skaičius tinklo apmokymui $M = 50000$, o taip pat iš anksto fiksuotas pradinių svorių rinkinys.

Buvo atlikti tyrimai su trimis realiomis keturmačių vektorių duomenų aibėmis atitinkamai su 28, 75 ir 150 vektoriais (Salinity, HBK ir Iris duomenų aibės). Iš pradžių

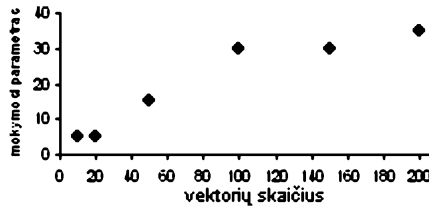
buvo nagrinėjama daugiamačių duomenų atvaizdavimo tikslumo priklausomybė nuo mokymosi parametro η , kai $\eta \in (0, 1)$. Gauti rezultatai rodo (2 pav.), kad didėjant mokymosi parametro reikšmei, gaunama geresnė vizualizavimo paklaida. Todėl, buvo atlikti bandymai su didesnėmis mokymosi parametro reikšmėmis, išeinant už intervalo $(0, 1)$ ribų. Rezultatai iliustruojami 3 pav. Pastebėta, kad geriausi rezultatai gauti, kai $\eta > 1$. Iš 2 pav. ir 3 pav. matyti, kad mokymosi parametro optimali reikšmė nagrinėjamiems duomenų aibėms yra intervale $[10, 30]$. Salinity ir HBK duomenų aibėms optimali mokymosi parametro reikšmė yra $\eta = 10$, Iris duomenų aibei – $\eta = 30$. Esant tokioms mokymosi parametro reikšmėms, gaunami geri vizualizavimo rezultatai, duomenys vizualizuojami greičiau ir tiksliau. Esant fiksuotam iteracijų skaičiui, geri vizualizavimo rezultatai pasiekiami per trumpesnę laiką, negu imant mokymosi parametro reikšmes iš intervalo $(0, 1)$.



2 pav. (a – Salinity, b – HBK, c – Iris) Vizualizavimo paklaidos priklausomybė nuo mokymosi parametro reikšmės η , $\eta \in (0, 1)$.



3 pav. (a – Salinity, b – HBK, c – Iris) Vizualizavimo paklaidos priklausomybė nuo mokymosi parametro reikšmės η , $\eta \in [1, 100]$.



4 pav. Optimalios mokymosi parametro reikšmės priklausomybė nuo duomenų aibės vektorių.

Atliekant tyrimus, taip pat buvo fiksuojamas algoritmo vykdymo laikas ir gautos vizualizavimo paklaidos kiekvienos iteracijos metu. Eksperimentai parodė, kad kuo didesnė mokymosi parametro reikšmė, tuo greičiau pavyksta pasiekti gerus vizualizavimo rezultatus. Tačiau didėjant mokymosi parametro reikšmei, didėja paklaidos svyravimai, ir tai gali sukelti tam tikrų SAMANN tinklo mokymo problemų.

Taip pat buvo atlikti eksperimentai su duomenų aibe, sudaryta iš atsitiktinai sugeneruotų vektorių. Tikslas buvo nustatyti, kokia turi būti optimali mokymosi parametro reikšmė ir kaip ji keičiasi didėjant duomenų aibės vektorių skaičiui. Duomenų aibės sudarytos atitinkamai iš 10, 20, 50, 100, 150 ir 200 keturmačių vektorių.

Duomenų aibės vizualizavimui buvo naudojami SAMANN tinklo parametrai: iteracijų skaičius $M = 7000$; mokymosi parametras $\eta \in [1, 50]$; paslėpto sluoksnio neuronų skaičius $n_2 = 20$. 4 pav. parodyti gauti rezultatai. Skirtingoms duomenų aibėms buvo gautos skirtingos mokymosi parametro reikšmės, priklausomai nuo aibės vektorių skaičiaus. Optimali mokymosi parametro reikšmė yra intervale $[5, 30]$.

Išvados

Straipsnyje ištirtas SAMANN tinklas, realizuojantis Sammono projekciją neuronų junginiu. Buvo nagrinėjama tinklo mokymo priklausomybė nuo mokymosi parametro η . Eksperimentai parodė, kad optimali mokymosi parametro reikšmė yra intervale $[5, 30]$. Pasirenkant tokias mokymosi parametro reikšmes galima žymiai sumažinti skaičiavimų trukmę ir gauti gerus vizualizavimo rezultatus per trumpesnę laiką, esant fiksuotam iteracijų skaičiui. Mažos mokymosi parametro reikšmės intervale $(0, 1)$ garantuoja stabilų mokymo paklaidos mažėjimą didėjant iteracijų skaičiui. Tuo tarpu, kai mokymosi parametro reikšmė pasirenkama didesnė, pastebimi tam tikri paklaidos svyravimai. Tačiau šie svyravimai yra pakankamai maži, kai mokymosi parametras pasirenkamas iš intervalo $[5, 30]$.

Literatūra

1. B. Lerner, H. Gutterman, M. Aladjem, I. Dinstein, Y. Romem, Feature extraction by neural network nonlinear mapping for pattern classification, in: *Conference on Pattern Recognition*, Track D, Vienna, Austria (1996), pp. 320–324.
2. A.K. Jain, R.C. Dubes, *Algorithms for Clustering Data*, Prentice-Hall (1988).
3. A.K. Jain, R. Duin, J. Mao, Statistical pattern recognition: a review, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **22**(1), 4–37 (2000).

4. J.J. Sammon, A nonlinear mapping for data structure analysis, *IEEE Trans. Computer*, **C-18**(5), 401–409 (1969).
5. A.K. Jain, J. Mao, Artificial neural network for nonlinear projection of multivariate data, in: *Neural Networks. IJCNN., International Joint Conference*, Vol. 3, 7–11 June (1992), pp. 335–340.
6. J. Mao, A.K. Jain, Artificial neural networks for feature extraction and multivariate data projection, *IEEE Trans. Neural Networks*, **6**, 296–317 (1995).
7. D. de Ridder, R.P.W. Duin, Sammon's mapping using neural networks: a comparison, *Pattern Recognition Letters*, **18**, 1307–1316 (1997).
8. R.A. Fisher, The use of multiple measurements in taxonomic problem, Part II, *Annual Eugenics*, **7**, 179–188 (1936).
9. D.M. Hawkins, D. Bradu, G.V. Kass, Location of several outliers in multiple regression data using elemental sets, *Technometrics*, **26**, 197–208 (1984).
10. D. Ruppert, R.J. Carroll, Trimmed least squares estimation in the linear model, *Journal of the American Statistical Association*, **75**, 828–838 (1980).

SUMMARY

V. Medvedev, G. Dzemyda. Optimization of the learning rate in the algorithm for data visualization

In this paper, we discuss the visualization of multidimensional data. A well-known procedure for mapping data from a high-dimensional space onto a lower-dimensional one is Sammon's mapping. The paper describes an unsupervised backpropagation algorithm to train a multilayer feed-forward neural network (SAMANN) to perform the Sammon's nonlinear projection. In our research the emphasis is put on the optimization of the learning rate to save computation time without losing the mapping quality.

Keywords: SAMANN network, visualization, learning rate, Sammon's mapping.