ზოგიერთი პარამეტრის გავლენა ხელოვნური ნეირონული ქსელის სწავლების სიზუსტეზე

მიხეილ კოტიშაძე
აკაკი წერეთლის სახელმწიფო უნივერსიტეტი, ქუთაისი, საქართველო
mmmkot56@mail.ru

ანოტაცია - სტატიაში განხილულია შემავალი პარამეტრების ნორმალიზაციის გავლენა ხელოვნური ნეირონული ქსელის სიზუსტეზე. ჩატარებულია ექსპერიმენტული კვლევა კომპიუტერულ მოდელზე, რომელშიც გამოყენებულია შეცდომის უკუგავრცელების ალგორითმი. ნაჩვენებია სხვადასხვა ტიპის ნორმალიზაციის გავლენის ხასიათი ქსელის სიზუსტეზე.

საკვანძო სიტყვები - ხელოვნური ინტელექტი, ნეირონული ქსელები, ნორმალიზაცია.

I. შესავალი

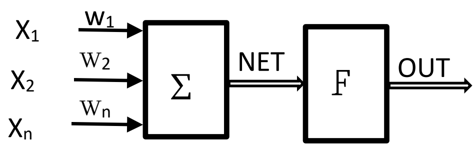
ნეირონული ქსელების გამოყენება საჭიროა სხვადასხვა ტიპის ამოცანების გადაწყვეტისას სხვადასხვა გარემოში. სხვადასხვა გარემო კი მოითხოვს სწავლების სხვადასხვა სიჩქარეს და სიზუსტეს. ნეირონული ქსელების სიჩქარე და სიზუსტე დამოკიდებულია სხვადასხვა პარამეტრებზე [2-4]. ამასთან სწავლების სიჩქარის გაზრდის მიზნით ხშირად მიმართავენ შემავალი პარამეტრების სხვადასხვა ტიპის ნორმალიზაციას [1]. ნორმალიზაციის ფორმა დამოკიდებულია გადასაჭრელი ამოცანის ტიპზე და იგი აირჩევა ინდივიდუალურად. ამასთან, სწავლების სიჩქარესა და სიზუსტეზე გარკვეულ გავლენას ახდენს ქსელის კონფიგურაციაც [4].

უნდა ითქვას, რომ არ არსებობს დადგენილი სტანდარტები და ერთიანი მიდგომა ნეირონული ქსელის შემავალი მონაცემების ნორმალიზაციის შესახებ.

აქედან გამომდინარე, აქტუალურია ამოცანა შემავალი პარამეტრების ნორმალიზაციის გავლენის კვლევა ხელოვნური ნეირონული ქსელის სიზუსტეზე.

II. ხელოვნური ნეირონული ქსელი

ნეირონული ქსელი, რომელზედაც ტარდებოდა ექსპერიმენტები, წარმოადგენს მრავალშრიან პერსეპტრონის კომპიუტერულ მოდელს შეცდომის უკუგავრცელების ალგორითმით [1]. ერთი ნეირონის სქემა, მოცემულია ნახ. 1-ზე.



ნახ. 1. ხელოვნური ნეირონის სქემა

აქ x_1, x_2, . . . x_n - ნეირონის შემავალი სიდიდეებია, რომლებიც აღწერენ შემავალ სახეებს; w_1, w_2, . . . w_n - კავშირების წონებია, რომლებსაც სწავლების საწყის ეტაპზე ენიჭებათ მცირე შემთხვევითი სიდიდეები (0-დან 1-მდე); **F(NET)** აქტივაციური ფუნქციაა და ყველაზე ხშირად აიღება კლასიკური სიგმოიდა. ეს ფუნქცია მოსახერხებელია, რადგანაც მას ძალიან მარტივი წარმოებული აქვს, რაც გამოყენებულია უკუგავრცელების ალგორითმის რეალიზაციისას.

სიგმოიდი, რომელსაც ზოგჯერ ლოგისტიკურ ან შემკუმშავ ფუნქციას უწოდებენ, ისე კუმშავს OUT-ს, რომ მისი მნიშვნელობა ძევს ნულსა და ერთს შორის. მრავალშრიან ნეირონულ ქსელებს მხოლოდ არაწრფივობის შემთხვევაში გააჩნიათ უფრო დიდი წარმოდგენითი სიმძლავრე, ვიდრე ერთშრიანებს. შემკუმშავი ფუნქცია უზრუნველყოფს საჭირო არაწრფივობას [1]

$$\mathrm{NET} = \sum_{i=1}^{n}(x_i\, w_i), \text{ ხოლო } \mathrm{OUT} = \frac{1}{1+e^{-\alpha \mathrm{NET}}} \qquad (1)$$

α - სიგმოიდას პარამეტრია.

ასეთი ნეირონებისაგან შექმნილია ქსელი რამოდენიმე შრით, თითოეულ შრეში ნეირონების განსაზღვრული რაოდენობით. შრის ყველა ნეირონის გამომავალი სიდიდე წარმოადგენს შემდეგი შრის ყველა ნეირონის შემავალ სიდიდეს (სრულკავშირიანი ქსელი). ნეირონების პირველი (შესავლებთან შეერთებული) შრე მხოლოდ სიგნალების გამანაწილებელ წერტილებს

წარმოადგენს და შემავალი სიგნალი უბრალოდ გადის მათში, მათსავე გამოსავალზე არსებულ წონებთან. ხოლო შემდეგი შრეების თითოეული ნეირონი იძლევა NET და OUT სიგნალებს (განტოლება 1). ქსელის სწავლების მიზანს წარმოადგენს წონების ისეთი შეწყობა, რომ შემავალი სიგნალების რაიმე სიმრავლემ მოგვცეს გამომავალი სიგნალების საჭირო სიმრავლე. შემავალი და გამომავალი სიგნალების სიმრავლეებს ვუწოდოთ ვექტორები. სწავლებისას იგულისხმება, რომ თითოეული შემავალი ვექტორისთვის არსებობს მისი მეწყვილე მიზნობრივი ვექტორი, რომლითაც მოცემულია მოთხოვნილი გამოსავალი. ორივეს ერთად ეწოდება შემსწავლელი წყვილი. როგორც წესი, ქსელს ასწავლიან მრავალ წყვილზე.

უკუგავრცელების ქსელის სწავლება მოითხოვს შემდეგი ოპერაციების შესრულებას:

1. შემსწავლელი სიმრავლიდან შევარჩიოთ მორიგი შემსწავლელი წყვილი; მივაწოდოთ შემავალი ვექტორი ქსელის შესავალზე.
2. გამოვთვალოთ ქსელის გამოსავალი.
3. გამოვთვალოთ სხვაობა ქსელის გამოსავალსა და მოთხოვნილ გამოსავალს (შემსწავლელი წყვილის მიზნობრივი ვექტორი) შორის.
4. მოვახდინოთ წონების კორექტირება ისე, რომ მოხდეს შეცდომის მინიმიზირება.
5. გავიმეოროთ ბიჯები 1-დან 4-მდე შემსწავლელი სიმრავლის თითოეული ვექტორისთვის მანამ, სანამ შეცდომა მთელ სიმრავლეზე არ მიაღწევს მისაღებ მნიშვნელობას.

1 და 2 ბიჯებით შესრულებული ოპერაციები იმ ბიჯებს ჰგავს, რომლებიც სრულდება უკვე ნასწავლი ქსელის ფუნქციონირებისას, ანუ მიეწოდება შემავალი ვექტორი და გამოითვლება მიღებული გამოსავალი. გამოთვლები სრულდება შრეობრივად. ჯერ გამოითვლება რაიმე j შრის ნეირონების გამოსავლები, შემდეგ ისინი გამოიყენება როგორც შემდეგი k შრის ნეირონების შესავლები, ხოლო k შრის ნეირონების გამოსავლები შეადგენს ქსელის გამომავალ ვექტორს.

ბიჯ 3-ზე ქსელის გამოსავლებიდან თითოეული, რომლებიც აღნიშნულია OUT-ით, გამოაკლდება მიზნობრივი ვექტორის შესაბამის კომპონენტს, რომ მივიღოთ შეცდომა. ეს შეცდომა გამოიყენება ბიჯ 4-ზე ქსელის წონების კორექციისთვის.

ამ ოთხი ბიჯის საკმაო რაოდენობით განმეორების შემდეგ ქსელის რეალურ გამოსავლებსა და მიზნობრივ ვექტორს შორის სხვაობა უნდა შემცირდეს მისაღებ სიდიდემდე. ამ დროს ამბობენ, რომ ქსელმა ისწავლა. შემდგომში ქსელი გამოიყენება სახეების გამოსაცნობად და წონები არ იცვლება.

1 და 2 ბიჯებზე ხორციელდება ”წინსვლა”, რადგანაც სიგნალი ვრცელდება შესავალიდან გამოსავალისკენ. ბიჯები 3 და 4 წარმოდგენენ ”უკუსვლას”, რადგანაც გამოთვლილი შეცდომის სიგნალი ვრცელდება ქსელში უკუმიმართულებით და გამოიყენება წონების კორექტირებისათვის. განვიხილოთ ორივე სვლა და წარმოვადგინოთ მათემატიკური ფორმით.

წინსვლა. ბიჯები 1 და 2 შეიძლება გამოვსახოთ ვექტორული ფორმით შემდეგნაირად: მიეწოდება შემავალი ვექტორი X (x_1, x_2, … x_n) და გამოსავალზე მიიღება ვექტორი OUT (OUT_1, OUT_2, OUT_3, . . . OUT_n,). ვექტორული წყვილი შესავალი-მიზანი X და T აიღება შემსწავლელი სიმრავლიდან. გამოთვლები წარმოებს ვექტორ X-ზე და მიიღება გამომავალი ვექტორი OUT.

როცა მიღებული იქნება გამოსავლების სიმრავლე, იგი წარმოადგენს შემავალ სიგნალებს შემდეგი შრის ნეირონებისათვის. პროცესი მეორდება ყველა შრეში მანამ, სანამ არ მიიღება ქსელის საბოლოო გამოსავლების სიმრავლე.

ერთი შრის გამომავალი ვექტორი წარმოადგენს შემავალ ვექტორს შემდეგი შრისთვის.

უკუსვლა. *გამომავალი შრის წონების შეწყობა.* რადგანაც გამომავალი შრის თითოეული ნეირონისთვის მოცემულია მიზნობრივი მნიშვნელობა, წონების შეწყობა ადვილად ხორციელდება მოდიფიცირებული დელტა-წესის გამოყენებით. შიგა შრეებს ”დაფარული შრეები” ეწოდებათ. მათი გამოსავლებისთვის არ არსებობს მიზნობრივი მნიშვნელობები შედარებისთვის, ამიტომ სწავლება რთულდება.

განვიხილოთ სწავლების პროცესი ერთი წონისათვის დაფარული j შრის p ნეირონიდან გამომავალი k შრის q ნეირონამდე. გამომავალი k

შრის ნეირონის გამოსავალი, გამოაკლდება რა მიზნობრივ T მნიშვნელობას, გვაძლევს შეცდომის სიგნალს. ის მრავლდება შემკუმშავი ფუნქციის წარმოებულზე α[OUT(1 – OUT), რომელიც გამოთვლილია k შრის ამ ნეირონისთვის და გვაძლევს δ სიდიდეს.

$$\boldsymbol{\delta} = \alpha\mathbf{OUT}(\mathbf{1} - \mathbf{OUT})(\mathbf{T} - \mathbf{OUT}) \qquad (2)$$

შემდეგ δ მრავლდება j შრის იმ ნეირონის OUT სიდიდეზე, რომლიდანაც გამოდის განსახილველი წონა. ეს ნამრავლი თავის მხრივ მრავლდება სწავლების სიჩქარის η კოეფიციენტზე (ჩვეულებრივ 0,01-დან 1,0-მდე) და შედეგი ემატება წონას. ასეთი პროცედურა სრულდება თითოეული წონისთვის დაფარული შრის ნეირონიდან, გამომავალი შრის ნეირონამდე.

საბოლოოდ, გამომავალი შრის ნეირონების წონების კორექციისთვის სრულდება შემდეგი გამოთვლები

$$\boldsymbol{\Delta w_{pq}} = \boldsymbol{\eta\delta_q}\mathbf{OUT_{pj}} \qquad (3)$$

$$\boldsymbol{w_{pq}(n+1)} = \boldsymbol{w_{pq}}(\mathbf{n}) + \boldsymbol{\Delta}\mathbf{w}_{\boldsymbol{pq}} \qquad (4)$$

სადაც $\boldsymbol{w_{pq}}(\mathbf{n})$ - დაფარული შრის p ნეირონის წონაა გამომავალი შრის q ნეირონამდე n ბიჯზე (კორექციამდე); $\boldsymbol{w_{pq}}(\mathbf{n+1})$ - წონის სიდიდეა n + 1 ბიჯზე (კორექციის შემდეგ);

წონების შეწყობა დაფარულ შრეში. განვიხილოთ გამომავალი შრის წინ მდებარე დაფარული შრის ერთი ნეირონი. წინსვლის დროს ეს ნეირონი თავის გამომავალ სიგნალს, შემაერთებელი წონის გავლით, გადასცემს გამომავალი შრის ნეირონებს. სწავლებისას ეს წონები ფუნქციონირებენ უკუ მიმართულებით, ანუ სიდიდე δ-ს გაატარებენ გამომავალი შრიდან დაფარული შრისკენ. ამ წონებიდან თითოეული მრავლდება გამომავალი შრის იმ ნეირონის δ-ზე, რომელთანაც ეს წონაა შეერთებული. დაფარული შრის ნეირონისთვის აუცილებელი δ სიდიდე მიიღება ყველა ასეთი ნამრავლის ჯამის გამრავლებით შემკუმშავი ფუნქციის წარმოებულზე:

$$\boldsymbol{\delta}_{\mathbf{q}} = \mathbf{OUT_{p,j}}(\mathbf{1} - \mathbf{OUT_{p,j}}) \sum \boldsymbol{\delta_q}\, \boldsymbol{w}_{\mathbf{pq}} \qquad (5)$$

როცა δ-ს მნიშვნელობა მიღებულია, შეიძლება კორექტირება გავუკეთოთ პირველი დაფარული შრის მკვებავ წონებს ზემოთ მოცემული ფორმულების საშუალებით, სადაც ინდექსები იცვლება შრეების შესაბამისად.

III. ქსელის ექსპერიმენტული კვლევა

კვლევა ჩატარდა ნეირონულ ქსელზე შეცდომის უკუგავრცელების კლასიკური ალგორითმით [1]. მოცემული იყო რამოდენიმე ამოცანა: მთელი რიცხვების შეკრება, მთელი რიცხვების გამრავლება, კვადრატული ფესვის ამოღება და სხვა. თითოეული სასწავლო ნაკრები შედგებოდა 50 შემსწავლელი წყვილისაგან. შემავალ შრეში 10 ნეირონია, ხოლო გამომავალ შრეში ერთი. რაც შეეხება ქსელის სხვა პარამეტრებს, ისინი წარმოდგენილია შემდეგი სახით:

- გამოყენებული სასწავლო წყვილები < 90%
- სწავლების სიჩქარის კოეფიციენტი - 0,1
- წყვილი გამოცნობილია თუ შეცდომა < 0,05
- ეპოქების (სასწავლო ნაკრების 1 სრული გავლა ქსელში) რაოდენობა - 50000
- აქტივაციური ფუნქცია - კლასიკური სიგმოიდა.
- სიგმოიდის პარამეტრი a - იცვლებოდა 0,4 - დან 1,8 - მდე

შეფასების კრიტერიუმად აღებული იყო :

- სწავლების პროცესში ქსელის მაქსიმალური შეცდომა
- სწავლების პროცესში ქსელის შეცდომის საშუალო.

კვლევისას შემავალი პარამეტრები ქსელს მიეწოდებოდა სამი სახით:

1. ნორმალიზაციის გარეშე
2. ნორმალიზაციით : (x – min)/(max – min)
3. ნორმალიზაციით : 1/(1+exp(-ax))

მიღებული შედეგები მოცემულია ცხრილებში.

ცხრილი 1. ნორმალიზაციის გარეშე

a	შეცდომის საშუალო	მაქსიმალური შეცდომა
0,4	0,0000425	0,0003563
0,6	0,00002823	0,0002582
0,8	0,000018965	0,0001814
1	0,00001598	0,0001567
1,2	0,000013401	0,0001274
1,4	0,00001297	0,0001277
1,6	0,00001527	0,00008501
1,8	0,00001251	0,00006725

ცხრილი 2. ნორმალიზაციით: (x – min)/(max – min)

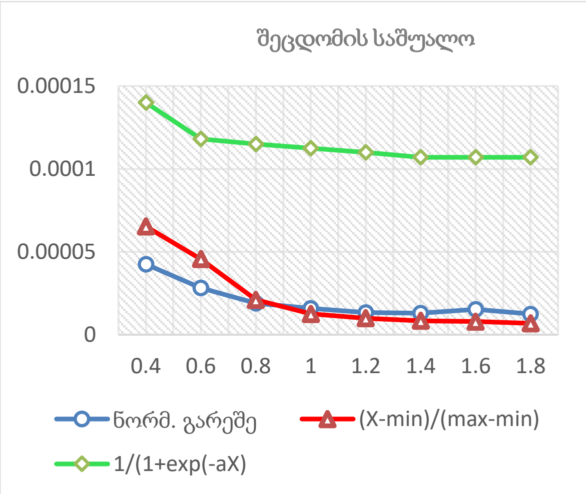
a	შეცდომის საშუალო	მაქსიმალური შეცდომა
0,4	0,00006533	0,00031
0,6	0,00004568	0,0002821
0,8	0,00002103	0,0001741

1	0,00001252	0,00012
1,2	0,000009915	0,00009686
1,4	0,000008388	0,00007539
1,6	0,000007946	0,00007269
1,8	0,000006863	0,00005965

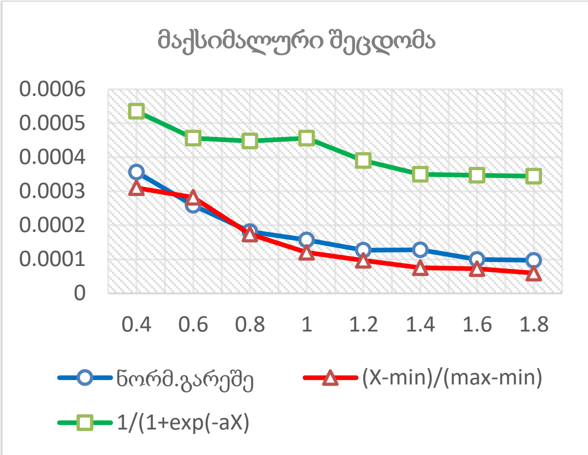
ცხრილი 3. ნორმალიზაციით : 1/(1+exp(-ax))

a	შეცდომის საშუალო	მაქსიმალური შეცდომა
0,4	0,00014	0,000535
0,6	0,000118	0,000456
0,8	0,000115	0,000448
1	0,0001125	0,0004558
1,2	0,00011	0,00039
1,4	0,000107	0,0003496
1,6	0,000107	0,0003473
1,8	0,000107	0,0003444

ექსპერიმენტების შედეგებით აგებული გრაფიკები გვიჩვენებენ, რომ როცა სიგმოიდის კოეფიციენტი ერთის ტოლია ქსელის შეცდომის საშუალო უახლოვდება მინიმუმს (ნახ.1) და მისი შემდგომი ზრდა ფაქტიურად არ ამცირებს შეცდომის საშუალო მნიშვნელობას. ამასთან, 1/(1+exp(-ax)) ნორმალიზაციით მიღებული სიზუსტე ბევრად უფრო დაბალია, ვიდრე ნორმალიზაციის გარეშე და (x – min)/(max – min) ნორმალიზაციით, რომელთა შედეგები დაახლოებით თანაბარია, თუმცა შეინიშნება (x – min)/(max – min) ნორმალიზაციის მცირე უპირატესობა.
რაც შეეხება მაქსიმალურ შეცდომას (ნახ.2), აქაც ანალოგიური სურათია.



ნახ. 1. ექსპერიმენტის შედეგები: ქსელის შეცდომის საშუალო.



ნახ. 2. ექსპერიმენტის შედეგები: ქსელის მაქსიმალური შეცდომა.

IV. დასკვნა

1. როცა სიგმოიდის კოეფიციენტი ერთის ტოლია ქსელის შეცდომის საშუალო უახლოვდება მინიმუმს და კოეფიციენტის შემდგომი ზრდა თითქმის არ მოქმედებს ქსელის სიზუსტეზე.
2. 1/(1+exp(-ax)) ნორმალიზაციით მიღებული ქსელის სიზუსტე ბევრად უფრო დაბალია, ვიდრე ნორმალიზაციის გარეშე და (x – min)/(max – min) ნორმალიზაციით, რომელთა შედეგები დაახლოებით თანაბარია, თუმცა მაინც შეინიშნება (x – min)/(max – min) ნორმალიზაციის მცირე უპირატესობა.
3. ქსელის მაქსიმალური სიზუსტის კანონზომიერება იგივეა, რაც ქსელის შეცდომის საშუალოსი.

ლიტერატურა

[1]. Ф. Уоссермен. Нейрокомпьютерная техника. М: Мир, 1992 г.

[2]. Царегородцев В.Г. Общая неэффективность использования суммарного градиента выборки при обучении нейронной сети // Материалы XII Всеросс. семинара " Нейроинформатика и ее приложения", Красноярск, 2004. - 196с. - С.145-151.

[3]. Царегородцев В.Г. Оптимизация экспертов boosting-коллектива по их кривым обучения // Материалы XII Всеросс. семинара "Нейроинформатика и ее приложения", Красноярск, 2004. - 196с. - С.152-157.

[4]. მ.კოტიშაძე. ხელოვნური ნეირონული ქსელის ზოგიერთი პარამეტრის კვლევა. VII საერთაშორისო სამეცნიერო-პრაქტიკული კონფერენცია „ინტერნეტი და საზოგადოება“ (INSO - 2013), ქუთაისი, 2015

Влияние некоторых параметров на точность обучения искусственных неиронных сетей

Михеил Котишадзе

Государственный университет им. Акакия Церетели

Аннотация

В статье рассмотрен вопрос влияния нормализации входных параметров на точность обучения искусственных неиронных сетей. Проведено экспериментальное иследование на компьютерном моделе сети, в котором исползован алгоритм обратного распространения ошибки. Показан характер влияния некоторих типов нормализации на точность обучения неиронных сетей.

The influence of some parameters on the accuracy of training artificial neural networks

Mikheil Kotishadze
Akaki Tsereteli State University

Abstrakt

The article considers the influence of normalization of input parameters on the accuracy of training artificial neural networks. An experimental investigation was carried out on the computer model of the network, in which the algorithm for back propagation of the error is used. The nature of the influence of some types of normalization on the accuracy of training artificial neural networks is shown.