ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი

თამარ ბლიაძე

ნეირონული ქსელების გამოყენება სპამის ფილტრაციასა და ანტი ვირუსულ პროგრამებში

ინფორმაციული სისტემები

სამაგისტრო ნაშრომი შესრულებულია ინფორმაციული სისტემების მაგისტრის აკადემიური ხარისხის მოსაპოვებლად

სამაგისტრო ნაშრომის ხელმძღვანელი: გელა ბესიაშვილი
(ტექნიკურ მეცნიერებათა კანდიდატი, ასისტენტ პროფესორი)

თბილისი, 2014

სარჩევი

შესავალი

რა არის „სპამ“ (Spam)? რატომ გვჭირდება მასთან ბრძოლა? შეგვიძლია თუ არა მისგან და მსგავსი პროდუქტებისგან საბოლოოდ თავის დაღწევა?

ოცდამეერთე საუკუნეში ელექტრონული ფოსტით სარგებლობა კომუნიკაციის გავრცელებული მეთოდია. პირველად სიტყვა e-mail გამოიყენეს 1978 წელს University of Medicine and Dentistry of New Jersey (UMDNJ)-ში სისტემისთვის, რომელიც იყო პირველი სრულყოფილი შიდაორგანიზაციული მეილ სისტემა. სულ რაღაც 20 წლის წინ email-ის მიღება ბევრად იშვიათი იყო ჩვეულებრივ ქაღალდის წერილთან შედარებით, როდესაც ეხლა პირიქითაა. ყოველდღიურად ვღებულობთ უამრავ მეილს სხვადასხვა კომპანიისგან თუ ნაცნობისგან, მაგრამ მათ შორის არსებობს ისეთებიც, რომელთა გამომგზავნს სულაც არ ვიცნობთ და არც მათი წერილების შინაარსი გვაინტერესებს დიდად. ასეთ შემთხვევებს, როდესაც mailbox-ში აღმოჩნდება ჩვენთვის არასასურველი და მომაბეზრებელი email, ვუწოდებთ სპამს. მათი რიცხვი დღითიდღე იზრდება და შესაბამისად საჭიროა მათგან დაცვის ხერხების შემუშავება. საშუალოდ ადამიანზე 10-15 სპამ მეილი მოდის და დაახლოებით 14.5 მიალრდია მათი დღიური რაოდენობა, რაც მთლიანი გაგაზავნილი მეილების 45 პროცენტია. კვლევებმა დაადგინა, რომ აშშ და სამხრეთ კორეა არიან ლიდერი სპამ გენერატორები. რა არის სპამის მიერ მიყენებული ზარალის ოდენობა? Radicati Research Group Inc. -ის მიხედვით ტიპიურ 1000 კაციან კომპანიას წელიწადში საშუალოდ 3 მილიონი აშშ დოლარის დახარჯვა უწევს მასთან საბრძოლველად (http://www.radicati.com/wp/wp-content/uploads/2010/04/Email-Statistics-Report-2010-2014-Executive-Summary2.pdf).

არსებობს არასასურველი მასობრივი მეილების (Unsolicited bulk email (UBE)) რამდენიმე სახეობა შინარსის მიხედვით, მაგალითად: “419” Scam, Phishing mail, ფარმაცევტული, რეპროდუქციების გასაღების, კაზინოები, წონის კორექციის, განათლების უფასოდ მიღების და სხვა. სტატისტიკურად ასე გამოიყურება ("Q1 2010 Internet Threats Trend Report" (PDF) (Press release). Commtouch Software Ltd. Retrieved 2010-09-23.):

Pharmacy	Replica	Enhancers	Phishing	Degrees	Casino	Weight Loss	Other
81%	5.40%	2.30%	2.30%	1.30%	1%	0,40%	6,30%

სპამი ანელებს ინტერნეტ ტრაფიკს, ტყუილ-უბრალოდ იყენებს სივცეს დისკზე, ამცირებს ქსელის გამტარუნარიანობას, რომ არაფერი ვთქვათ ადამიანის მიერ დახარჯულ დროზე მათი კლასიფიკაციისთვის. წყაროები, რომლებიც ხშირად ცნობილი კომპანიების სახელს არიან ამოფარებულნი, აგზავნიან Scam-ს და ცდილობენ მომხმარებლის არაკომპეტენტურობით ისარგებლონ და პლასტიკური ბარათის მონაცემები დასტყუონ.

არამარტო ამ და სხვა მიზეზების გამო, ყოველდღიურად მიღებულ არასასურველ მეილებთან/ინფორმაციასთან ბრძოლა და მათი აღმოჩენა დღესდღეისობით ფრიად აქტუალური თემაა, რასაც ემსახურება ეს კონკრეტული კვლევა.

ცნობილი კომპანიები დიდ ყურადღებას უთმობენ ამ თემაზე მუშაობას. ისინი, რათქმაუნდა, თავიანთ სპამ ფილტრის მუშაობის ალგორითმს, რომელიც კომპანიის კონფიდენციალური ინტელექტუალური საკუთრებაა, არ გასცემენ, მაგრამ მაგალითად Google ზოგადად აღწერს Gmail-ის სპამთან ბრძოლის რამდენიმე კრიტერიუმს: პირველ რიგში აყენებს Community Feedback-ს, როდესაც მრავალი ადამიანი მონიშნავს კონკრეტულ მეილს/გამომგზავნს სპამად, ის ავტომატურად გლობაურად ხდება ბლოკირების ობიექტი; ასევე, მომხმარებლის მიერ შექმნილი ფილტრები; დაბოლოს ავტომატური სპამ ფილტრი ().

წარსულში სპამთან ბრძოლისთვის პროგრამები იყენებდნენ keyword ფილტრებს: კონკრეტული სიტყვა თუ ფიგურირებდა სათაურში და/ან მეილის ტანში, მიიჩნეოდა სპამად. ეს მეთოდი მალე მოძველდა, რადგან საეჭვო სიტყვები მრავლდებოდნენ, სპამი თავად გახდა მრავალფეროვანი და მომხმარებლის დროის ეკონომია სულაც აღარ ხდებოდა. შემდეგში შემოღებული სპამ ფილტრები შეიძლება დაიყოს შემდეგნაირად:

მომხმარებლის მიერ შექმნილი ფილტრები: სათაურის და შიგთავსის მიხედვით გადაიტან სპამ ფოლდერში.

სათაურის ფილტრი: მხოლოდ მეილის სათაურის მიხედვით განისაზღვროს მისი სტატუსი.საკმაოდ არასანდოა.

ენობრივი ფილტრი: ისინი უბარლოდ ფილტრავენ მეილებს , რომლებიც არ არიან მშობლიურ ენაზე, თუმცა ასეთი რამ არაა ეფექტური, თუ ეს უცხოური ენა ინგლისურია .

უფლების დონეზე შეზღუდვა: ის უბრალოდ ბლოკავს ყველა მეილს რომელიც არაავტორიზებული წყაროდანაა მოსული. ამ შემთხვევაში გამგზავნს მისდის მეილი, რომელიც სთხოვს სპეციალურ ფორმაზე ინფორმაციის შევსებას.

შინაარსობრივი ფილტრი: ისინი არჩევენ მეილის ტექსტს ნეირონული ქსელისა თუ არამკაფიო ლოგიკის გამოყენებით, ასევე ბაიესის ალგორითმით, რათა დაადგინონ წონა ამ მეილისა - არის თუ არა სპამი. ისინი ეფექტურებია, თუმცა შეიძლება საჭირო წერილიც მიაკუთვნონ სპამს. ასევე სპამის ერთ-ერთი სახეობაა სურათიანი წერილი, სადაც ტექსტი სურათზეა გადატანილი.

ამ ნაშრომის მიზანია სწორედ ხელოვნური ნეირონული ქსელების როლის და წარმატებების კვლევა ამ კუთხით, რადგან აღწერილი მეთოდები მუდამ დასახვეწი იქნება სპამერთა რიცხვის და მათი მუშაობის ხერხების არათუ შემცირდების, არამედ მომვალში უფრო და უფრო გამრავალფეროვნების გამო.

ლიტერატურის მიმოხილვა

სპამის აღმოჩენა და მისი პრევენცია პოპულარული თემაა, რომელზეც მიმდინარეობს კვლევები და განიხილება სხვადასხვა მეთოდები. ის ოთხ კატეგორიადაა დაყოფილი :

1. ინდივიდის მიერ დაკლასიფიცირება,
2. ავტომატიზირება სისტემის ადმინისტრატორის მიერ,
3. ავტომატიზირება მეილის გამომგზავნის მიერ
4. კანონმდებლოით რეგულირება.

ამათგან ჩვენთვის საინტერესოა სისტემის ადმინისტრატორების მიერ გამოყენებული ანტი-სპამ სისტემების გამოკვლევა და მათი გაუმჯობესება. ეს პროგრამები ხშირად იყენებენ მანქანური დასწავლის მეთოდებს, რომელიც არის ხელოვნური ინტელექტის დარგი და გულისხმობს სისტემის არსებულ მონაცემებზე დასწავლას. განასხვავებენ რამდენიმე დასწავლის ალგორითმს:

1.დასწავლა მასწავლებლით 2. მასწავლებლის გარეშე 3. ნახევრად მასწავლებლით დასწავლა 3. ტრანსდუქცია 4. გაძლიერებული/მაქსიმიზირებული სწავლება 5. დასწავლა სწავლისთვის 6. განვითარებადი სწავლება (რობოტები) (wiki Machine learning).

დასწავლის მეთოდები მრავალია, მათ შორის კი გამოირჩევა: ბაიესის თეორემა გადაწყვეტილების მიღებისთვის (Naive Bayes classifier , Bayesian network), კლასტერიზაცია, გადაწყვეტილების მიღების ხე (Decision Trees), Support vector machines, ნეირონული ქსელები მრავალშრიანი პერცეპტრონით (Artificial neural networks MPL) [Introduction to Machine Learning http://www.realtechsupport.org/UB/MRIII/papers/MachineLearning/Alppaydin_MachineLearning_2010.pdf]

ყველზე ხშირად სპამის ფილტრაციისთვის გვხვდება ბაიესის და ნეირონული ქსელების მეთოდები. ასევე ვექტორული მანქანური დასწავლა და ხანდახან პეტრის ქსელიც.

ბაიესის მეთოდი ეყრდნობა ამავე სახელწოდების თეორემას

$$P(W|L) = \frac{P(L|W)P(W)}{P(L)} = \frac{P(L|W)P(W)}{P(L|W)P(W) + P(L|M)P(M)}$$

…(განვრცობა)

… (ჩასართავია) მნიშვნელოვანია იმ ფაქტის აღნიშვნა, რომ არსებობს რისკი საჭირო მეილის (ham) მონიშვნისა როგორც spam (False Positive), რაც ადამიანის მიერ ბევრად უფრო მტკივნეულად აღიქმება, ვიდრე არასასურველი წერილის Inbox-ში მოხვედრა (False Negative).

მეთოდები

... სამუშაო იყოფა 2 ძირითად ნაწილად: შემოსული ტექსტის ანალიზი და ამ ანალიზის შედეგად მიღებული ინფორმაციის საფუძველზე ნეირონულ ქსელში მისი გატარება და დადგენა ამ ტექსტის მქონე მეილის სტატუსი - არის თუ არა ის სპამი.
რადგან წერილების უმეტესობა ჯერ კიდევ ტექსტური სახით მოდის, მისი ანალიზი საჭირო კომპონენტია სპამის აღმოჩენის პრობლემისთვის.
ტექსტის ანალიზისთვის გამოიყენება რამდენიმე მხასიათებელი და სხვადასხვა საზომები, მაგალითად :

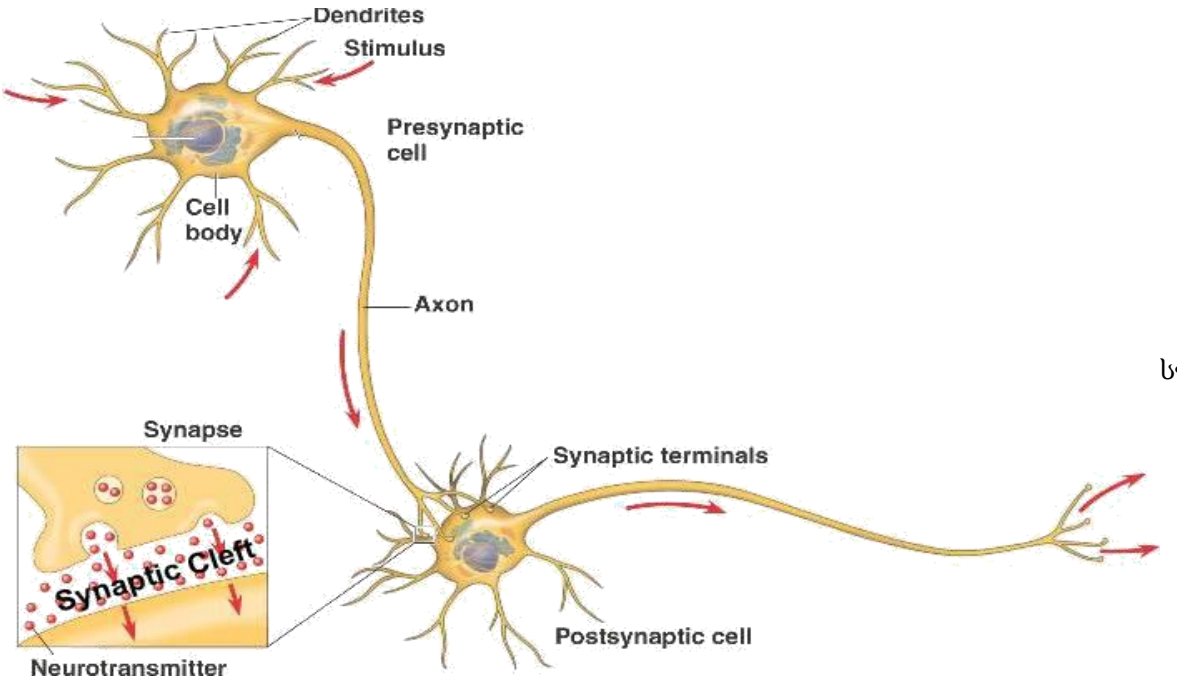
- ასოთა სრული რაოდენობა (აღვნიშნოთ როგორც C)
- მთავრულ ასოთა რაოდენობა / C , მთავრულთა თანაფადრობა
- ციფრთა რაიდენობა / C
- ცარიელ ადგილთა (space) რაოდენობა /C
- ასოების სიხშირე / C (კლავიატურაზე ანბანის ყველა ასო + ციფრები)
- special characters (*, _ ,+,=,%,$,@,- , \,/) -ის სიხშირე
- სიტყვათა რაოდენობა (აღვნიშნოთ როგორც M)
- მოკლე სიტყვათა რაოდენობა / M
- ასო-ბგერათა რაოდენობა სიტყვებში / C
- სიტყვათა საშუალო სიგრძე
- წინადადების საშუალო სიგრძე ასოების მიხედვით
- წინადადების საშუალო სიგრძე სიტყვების მიხედვით
- უნიკალური სიტყვების რაოდენობა / (M)
- Hapax Legomenon, მხოლოდ ერთხელ შეხვედრილი სიტყვების სიხშირე
- მხოლოდ ორჯერ შეხვედრილი სიტყვების სიხშირე.
- Yule's K სიდიდე (მსგავსი სიტყვების სიხშირის დადგენის სიდიდე)
- პუნქტუაციის ნიშნების სიხშრე (. ‹ ; ? ! : ( ) – “ « » < > [ ] { })

ეს მახასიათებლები, ამავე მიმდევრობით, აღწერენ კონკრეტულ მეილს და შესაძლებელი ხდება მატრიცის ვექტორების სახით გადაეცეს ნეირონულ ქსელს მთელი წყება მეილებისა ჯერ დასასწავლად, შემდეგ გასატესტად.

მეილების ბაზად აღებულია Enron Mail Dataset (http://www.csmining.org/index.php/enron-spam-datasets.html , http://www.cs.cmu.edu/~./enron/), რომელიც წარმოადგენს გაკოტრებული ამერიკული ფირმის Enron მეილების გასაჯაროებულ კრებულს. ასევე ხშირად სატესტოდ გამოიყენება Spambase, რომელიც წარმოადგენს მეილებიც მატრიცას, შედგენილს 58 ატრიბუტიგან (http://www.ics.uci.edu/~mlearn/MLRepository.html)

ნეირონული ქსელი

ნეირონული ქსელი (Artificial Neural Network) წარმოადგენს გენეტიკური ალგორითმების ერთ-ერთ განაყოფს, რომელიც მიმართულია ადამიანის ტვინის მსგავსი ხელოვნური ქსელის სიმულირების მოდელირებისკენ. ტვინი შედგება უამრავი (10^{11}) სპეციფიკური უჯრედისგან - ნეირონისგან. ისინი შედგებიან თავად უჯრედისგან (სომა), და მისი წანაზარდებისგან - დენდრიტი და აქსონი. ეს 2 უკანასკნელი წარმოადგენს სინაპტიკური მუხტების მიმღებ/გადამცემთ - დენრდიტი მიიღებს, ხოლო გრძელი აქსონი ნეირონის უჯრედიდან გამოიტანს და გადასცემს სხვა ნეირონს, მისი დენდრიტის გავლით (ან პირდაპირ სომაზეც). ანუ ტიპიური სინაფსი გამოიყურება შემდეგნაირად :



სურ. 1

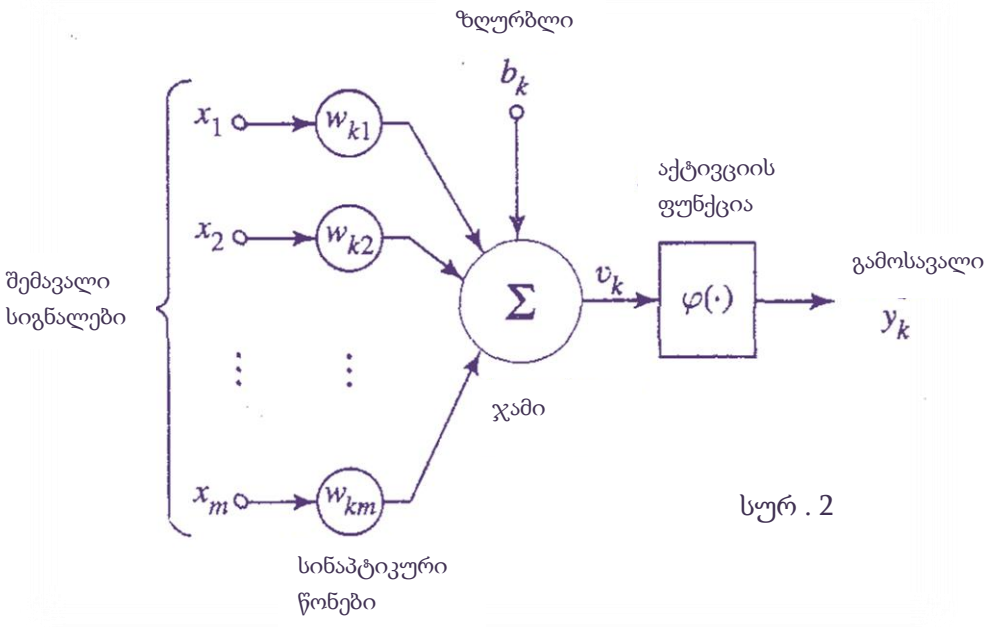
ტვინი შეიძლება მივიჩნიოთ ასეთი ნეირონების მრავალშრიან ციკლურ სტრუქტურად, სადაც გარე სამყაროს გამღიზიანებლებიდან (შემავალი ფენა) მიღებული იმპულსები გადაეცემა ტვინის ქერქს (შუალედურ ფენას), სადაც ხდება მათი დამუშავება და შემდეგ ისევ გარე სამყაროს დაუბრუნდება („გამოსავალი“ ფენაზე) სამოქმედო შედეგი.

მათემატიკური მოდელი

ხელოვნური ნეირონული ქსელი ახდენს ტვინის მოდელირებას, რომელიც აერთიანებს ხელოვნურ ნეირონებს. ასეთ მოდელს შეუძლია ინფორმაციის მიღება გარედან, მისი დამუშავება და შედეგის გამოტანა.

როგორც სურ. 1-დან ჩანს, ნეირონს შეიძლება გააჩნდეს მრავალი ინფორმაციის მიმღები (დენდრიტი) და ერთი გამომავალი წერტილი (აქსონი). ასევე, ხელოვნურ ნეირონში მათემატიკური კუთხით, ნეირონულ ქსელი ასრულებს რამდენიმე $x(1), x(2), .. x(n)$ შემავალი სიგნალების არაწრფივ გარდაქმნას y გამომავალ სიგნალად.

ინფორმაციის დამმუშავებელი ნეირონის სტრუტურა წარმოდგენილია სურ. 2-ზე:



სურ . 2

ის შედგება

- *შემავალი სიგნალები* ($x(1), x(2), .. x(n)$;

- *სინაპტიკური წონები* (w_{kj}), რომლებიც არის „სიმძლავრე“ სინაფსისა x_j შემავალი სიგნალიდან k ნეირონამდე. ტვინის ნეირონებისგან განსხვავებით, ხელოვნური ნეირონის წონა უარყოფითი მნიშვნელობისაც შეიძლება იყოს.
- *ჯამი* - რომლიც კრებს ყველა სიგნალის წონას;
- *აქტივაციის ფუნქცია,* რომელიც წარმოადგენს ნეირონის გამოსავლის ამპლიტუდის შემამცირებელს (ამიტომ მას შემამჭიდროვებელასც უწოდებენ ხოლმე). მიღებულია, რომ გამოსავლის დიაპაზონი გამოისახოს როგორც [0,1] ან [-1,1].

მოდელს ასევე გააჩნია დამატებითი სიდიდე - *ზღურბლი* , რომელიც ზრდის ან ამცირებს აქტივაციის ფუნქციის შემავალ მნიშვნელობებს, დამოკიდებულს იმაზე, ის დადებითია თუ უარყოფითი.

მათემატიკური მოდელით ნეირონი k გამოისახება შემდეგნაირად :

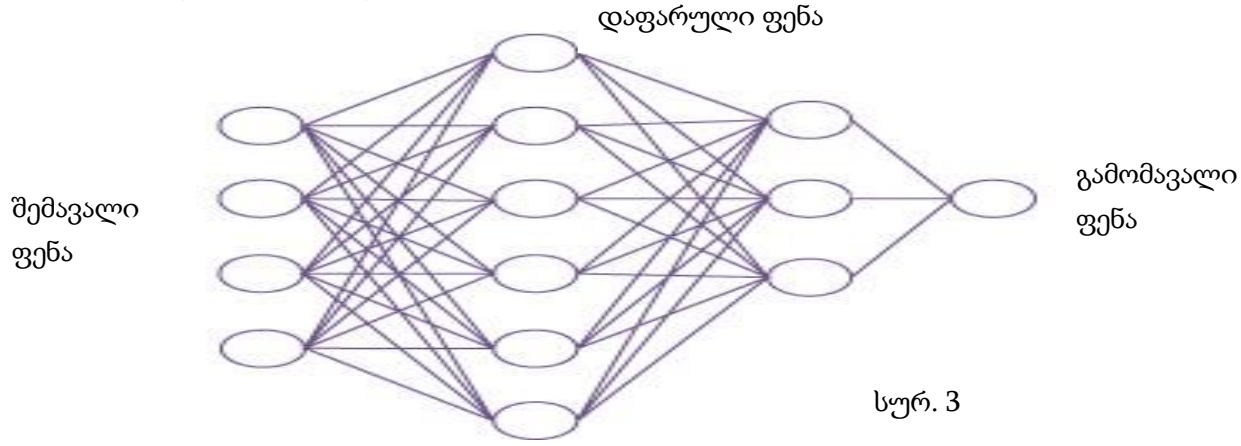
$$u_k = \sum_{j=1}^{m} w_{kj} x_j$$

და

$$y_k = f(u_k + b_k)$$

სადაც b_k არის ზღურბლი, y_k არის გამოსავალი სიგნალი და $f()$ არის აქტივაციის ფუნქცია.

აღნიშნული მოდელი აღწერს 1 ნეირონს, მაგრამ ქსელში რამდენიმე ნეირონისგან შემდგარი ფენა შეიძლება შეიქმნას, სადაც წინა ფენის გამომავალი სიგნალები შეიქმნებიან შემდეგი ფენის შემავალ მონაცემებად (სურ. 3);



სურ. 3

აქტივაციის ფუნქცია

აქტივაციის ფუნქციას ნეორონულ ქსელში გამოიყენებენ რამდენიმეგვარს:

- ზღურბლური (Threshold function) $f(u) = 1$, თუ $u \geq 0$ და $f(u) = 0$, თუ $u < 1$
- უბან-უბან წრფივი
    - $f(u) = 1$,თუ $u > 0{,}5$;
    - $f(u) = |u|$,თუ $|u| < 0{,}5$
    - $f(u) = 0$, თუ $u < 0{,}5$
- სიგმოიდური

$$f(u) = \frac{1}{1 + e^{-bu}}$$

ჩვენს შემთხვევაში ვირჩევთ აქტივაციის სიგმოიდურ ფუნქციას (რადგან ვირჩევთ მრავალშრიან პერცეპტრონს (Multi-Layer Perceptron))

სწავლება

ნეირონული ქსელი საინტერესო და განსხვავებული იმითაა, რომ მას შეუძლია დასწავლა, ისევე როგორც ნამდვილ, ცოცხალ ტვინს. ქსელის დასწავლა არსებობს მასწავლებლით (Supervised) და მასწავლებლის გარეშე (Unsupervised). მასწავლებლის გარეშე დასწავლა მიმდინარეობს ქსელის თვითორგანიზების საფუძვლეზე. მეილების ფილტრაციისთვის არჩეულია მასწავლებლით დასწავლის მეთოდი, რომელიც შემდეგ პრინციპზეა დაფუძნებული: ქსელის მეირ დასასწავლი მაგალითი შედგება 2 ნაწილისგან: შემავალი მონაცემები და სასურველი შედეგი, რომელსაც უნდა შედარდეს გამოსული შედეგი. აქვე აღინიშნოს, რომ სასურველ შედეგსა და მეღებულ შედეგს შორის სხვაობა არის შეცდომა (error), რომელიც უნდაგამოსწორდეს ქსელის მიერ. სწავლების პროცესი კი მის მინიმუმამდე დაყვანისას ხორციელდება.

მასწავლებლით დასწავლის ცნობილი ალგორითმია უკუგანვრცობის ალგორითმი (Backpropagation algorithm), რომელიც გამოსავალზე დასახული მიზნის მისაღწევად, გამოსულ შედეგს, თუკი ის არ ემთხვევა მას, ბრუნდება უკუმიმართულებით შემომავალი ფენისკენ და ახდენს error-ის კორექციას. სწორედ ამ მეთოდში გამოგვადგება ზემოთ აღნიშნული აქტივაციის სიგმოიდური ფუნქცია. საჭირო მონაცემები ამ მეთოდის განსახორციელებლად

1. E შესავალი სიგნალი, და შესაბამისი გამოსავალი C მნიშვნელობა.

2. გამოითვალოს E პირდაპირი განვრცობა (Feed Forward) ; ქსელში განისაზღვრება წინითი სიდიდეები S_I "და აქტივატორები U_I თითოეული უჯრისთვის.
3. გამოსასვლელიდან დაიწყება მოძრაობა შემავალი ფენისკენ, ფარული შრეების გავლით, ითვლის შედცომის მნიშვნელობებს
    a. $\delta_{0=(C_{I-X(i)})X(5)(1-X(i))}$ გამოსავალი უჯრისთვის
    b. $\delta_{i=(\sum_{mim>i} W_{MJ}\delta_0)}X(i)(1-x(i))$ ფარული უჯრებისთვის
4. ქსელში წონების შემდეგნაირად გადანაწილება

$$w^*_{IJ} = W_{IJ} + \rho\delta_0 X(i)$$

აღსანიშნავია ფაქტი , რომ შემავალი მონაცემების ვექტორები რაოდენობა, ასევე ამ ვექტორის ატრიბუტების რაოდენობა გავლენას ახდენს ქსელის მუშაობის სიწორეზე და სისწრაფეზე. თუკი პარამეტრები ბევრად მეტია შემავალ მონაცემების ერთობლიობაზე, დიდი ალბათობით მოხდება ე.წ. ზედმეტი დასწავლა (Overfitting). მისაღები თანაფარდობაა პარამეტრებზე ათჯერ მეტი შემავალი ვექტორები , ე.წ. არახმაურიანი გამოსავლისთვის (target), და დაახლოებით ორჯერ მეტი - ეს მინიმალური სასურველი მოთხოვნაა.

კვლევის შედეგები და მათი განხილვა

დასკვნა

გამოყენებული ლიტერატურის სია

1. Simon Haykin; Neural Networks – A Comprehensive Foundation ; Second Edition
2. Ethem Alpaydın - Introduction to Machine Learning; Second Edition; The MIT Press; Cambridge, Massachusetts; London, England
3. Email Statistical Report http://www.radicati.com/wp/wp-content/uploads/2010/04/Email-Statistics-Report-2010-2014-Executive-Summary2.pdf
4. Spam? Not Any More! Detecting Spam emails using neural networks. Thiagarajan Sivandayan. ECS/CS/ME 539 Project.
5. Anti-Spam Filtering Using Neural Networks and Bayesian Classifiers. Yue Yang and Sherif Elfayoumy. Proceedings ofthe 2007 IEEE International Symposium on Computational Intelligence in Robotics and Automation
6. Spam filtering - Alessandro Zanni and Israel Saeta Pérez, UPC ; January 2012
7. Support Vector Machines for Spam Categorization - Harris Drucker, Senior Member, IEEE, Donghui Wu, Student Member, IEEE, and Vladimir N. Vapnik
8. ..