

VLERËSIMI I ÇMIMIT TË PASURIVE TË PALUAJTSHME NËPËRMJET APLIKIMIT TË RRJETEVE NEURALE

ELDA XHUMARI¹, ARTUR BAXHAKU²

¹Universiteti i Tiranës, Fakulteti i Shkencave të Natyrës, Departamenti i
Informatikës

²Universiteti i Tiranës, Fakulteti i Shkencave të Natyrës, Departamenti i
Matematikës

e-mail: elda.xhumari@fshn.edu.al

Përmbledhje

Rrjetet neurale janë një instrument i fuqishëm që përdoren për zgjidhjen e shumë problematikave, duke u bazuar në parimin e funksionimit të sistemin nervor biologjik. Aftësia e rrjeteve neurale për t'u trajnuar dhe mësuar, si dhe për të korrigjuar gabimet, përbën një avantazh për zgjidhjen e problematikave në çdo fushë të jetës. Qëllimi i këtij punimi konsiston në vlerësimin e çmimit të pasurive të paluajtshme, nëpërmjet aplikimit të topologjive të ndryshme të rrjeteve neurale. Në këtë artikull, do të trajtohen disa modele vlerësimi që janë të përshtatshme për tregun e pasurive të paluajtshme.

Fjalëkyçe: Vlerësimi, seritë kohore, rrjetet neurale, arkitektura e rrjetit neural.

Abstract

Neural networks are a powerful tool used to solve many problems, based on the principle of functioning of the biological neural system. The ability of neural networks to train and learn, as well as to test and correct mistakes, is an advantage in solving problems in every area of life.

The purpose of this paper is to evaluate the price of real estate, through the application of different topologies of neural networks. In this article, we will address some valuation models that are suitable for the real estate market.

Keywords: Evaluation, time series, neural networks, neural network architecture.

Hyrje

Rrjetet neurale artificiale (Hubel & Wiesel, 2005) përbëjnë një instrument mjaft të fuqishëm në ditët e sotme për zgjidhjen e detyrave të ndryshme, ndër të detyrat e vlerësimit, klasifikimit apo parashikimit të të dhënave. Një nga përparësitë kryesore të rrjeteve neurale është mundësia për të përafëruar apriori funksionet jo-lineare, pa asnjë informacion paraprak në lidhje me veçoritë e të dhënave. Një tjetër përparësi e përdorimit të rrjeteve neurale për vlerësimin dhe parashikimin e të dhënave është aftësia për të mësuar vetëm nga shembujt; gjithashtu rrjetet neurale janë në gjendje të kapin varësi të fshehura dhe jo-lineare, edhe nëse ka një zhurmë të konsiderueshme në setin e trajnimit.

Si disavantazhe të rrjeteve neurale përmendim se ato mund të mësojnë varësinë, të vlefshme vetëm për një periudhë të caktuar kohe dhe gabimi i parashikimit mund të mos vlerësohet me saktësi. (Nielsen, 2015)

Qëllimi i këtij studimit është të aplikojë dhe krahasojë metoda të njohura për përcaktimin e një modeli në funksion të vlerësimit të ndryshimit të çmimeve në tregun e transaksioneve reale të pasurive të paluajtshme, bazuar në teknologjitë e rrjeteve neurale, të cilat do të mund të përmirësojnë ndjeshëm efikasitetin e agjencive të pasurive të paluajtshme.

Në fazën fillestare, do të analizohen metodat e njohura të vlerësimit të të dhënave, duke përdorur rrjetet neurale artificiale.

Faza e dytë konsiston në zhvillimin e një modeli dhe algoritmi të vlerësimit të attributeve që ndikojnë në çmimin e pasurive të paluajtshme.

Faza e tretë është studimi i algoritmit të parashikimit të çmimit të pasurive të paluajtshme. Për realizimin e detyrave të sipërpërmendura do të përdoren funksionalitetet e paketës MATLAB (Matlab & Simulink) si dhe metodat e teorisë së rrjeteve neurale artificiale.

Metodologjia e eksperimenteve

Në këtë paragraf paraqitet një zgjidhje e problemit të vlerësimit të çmimit të pasurive të paluajtshme në qytetin e Tiranës nëpërmjet realizimit të disa eksperimenteve. Janë përdorur të dhënat për objektet e pasurive të paluajtshme (apartamente) të marra gjatë periudhës nga 1 qershor 2019, deri më 1 nëntor 2019. Secili prej eksperimenteve mund të ndahet në disa faza.

Hapi i parë konsiston në formimin e një seti që do të përdoret për të trajnuar modelin. Në këtë fazë përcaktohet se si do të paraqiten të dhënat (neuronet hyrëse dhe dalëse).

Vlerësimi duke përdorur rrjetin neural artificial konsiston në të mësuarit e historikut të ndryshimit të variablilit në një kohë të kufizuar të zgjedhur dhe aplikimin e informacionit të mësuar për të ardhmen. Të dhënat nga e kaluara do të shërbejnë si inputet e rrjetit neural artificial dhe si output marrim çmimin e pasurive të paluajtshme në të ardhmen. Për realizimin e eksperimenteve është përdorur “të mësuarit me mësues” ose “të mësuarit me mbikëqyrës”, i cili konsiston në përfshirjen e ndryshimit të peshave të rrjetit nëpërmjet aplikimit të një grupi kampionësh trajnimit, i cili përfaqësohet nga inputi dhe outputi i dëshiruar. Ky trajnim përsëritet në mënyrë të vazhdueshme, për të arritur një gjendje të qëndrueshme të peshave sinaptike, të cilat ndryshohen për të minimizuar diferencën midis daljes së dëshiruar dhe asaj aktuale të rrjetit.

Për zgjidhjen e problemit të vlerësimit përdoren lloje të ndryshme rrjetesh neurale, si multilayer perceptrons (Du & Swamy, 2019), feed-forward neural networks (Annema, 2012), radial basis functions (Liu, 2013) convolutional neural networks (Michelucci, 2019), etj. Për realizimin e qëllimit të këtij punimi është zgjedhur modeli multilayer perceptron, duke konsideruar

aftësinë e tij gjithëpërfshirëse në trajtimin e një spektri të gjërë problemesh nga fusha të ndryshme.

Për trajnimin e modelit janë përdorur të dhëna të marra nga faqe të agjencive imobiliare, si: Remax, Century21 dhe Future Home, për periudhën e sipërpërmendur për qytetin e Tiranës.

Të dhënat e disponueshme janë ndarë në tre grupe: set trajnimi (i të mësuarit), set vlerësimi dhe set testimi. Këto sete mund të mbivendosen dhe nuk duhet të jenë të vazhdueshme vazhdueshme (James & Witten & Hastie & Tibshirani, 2013). Seti i të mësuarit është një sekuencë që i paraqitet rrjetit neural gjatë fazës së trajnimit (mësimi). Rrjeti përshtatet me të për të arritur rezultatet e kërkuara (me fjalë të tjera, pesha në rrjet ndryshohet bazuar në këtë set). Diferenca në daljen e kërkuar matet duke përdorur setin e vlerësimit dhe ky ndryshim përdoret për të vërtetuar nëse trajnimi i rrjetit mund të përfundojë. Seti i fundit, grupi i testimit, përdoret më pas për të provuar nëse rrjeti është në gjendje të punojë gjithashtu me të dhënat që nuk janë përdorur në procesin e mëparshëm.

Për ta përmbledhur, seti i trajnimit përdoret për krijimin e një modeli, seti i vlerësimit përdoret për verifikimin e modelit, dhe seti i testimit përdoret për testimin e përdorshmërisë së modelit.

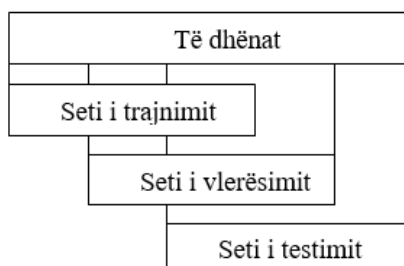


Figura 1. Setet e të dhënave

Përpunimi i të dhënave është një proces shumë i rëndësishëm, sidomos për rrjetet neurale ku outputi duhet të jetë brenda një diapazoni të caktuar vlerash. Pra, është e pamundur të parashikosh vlera jashtë këtij intervali dhe për këtë arsye nevojitet normalizimi i të dhënave. Të dhënat origjinale konvertohen paraprakisht në një format të atillë, që të mund të shërbejnë si input për rrjetin neural. Për shembull, kemi të dhëna sasiore që ndikojnë në çmimin e një apartamenti, si numri i kateve, numri i dhomave, sipërfaqja, dhe të dhëna cilësore si zona dhe tipi/materiali i ndërtimit. Të dhënat cilësore, nga bashkësia e të dhënave $x = \{x_1, x_2, \dots, x_n\}$, $n = \overline{1, i}$ ku i - numri i elementëve të setit x (për shembullin tonë $i = 300$), duhet të kuantikohen dhe normalizohen. Për kriterin cilësor “zona”, janë marrë në konsideratë njësitë administrative të Tiranës, ku çdo njësi administrative është paraqitur sipas numrit përkatës, për shembull Njësia administrative 1, ka numrin një.

Për kriterin “tipi/materiali i ndërtimit”, kuantifikimi ose paraqitja me vlerë numerike është bërë si në tabelën Nr 1 (Paraqitja me vlerë numerike e materialeve të ndërtimit):

Tabela 1. Paraqitja me vlerë numerike e materialeve të ndërtimit

Materiali	Numri
Tulla	1
Beton	2
Gur	3
Dru	4
Parafabrikate	5
Konstruksion metalik	6

Normalizimi i tyre është realizuar duke përdorur formulën e mëposhtëme:

$$x_{i,norm} = \frac{x_i}{x_{max}}$$

Ku x_{max} - vlera maksimale e x dhe $x_{i,norm}$ - vlera e normalizuar e x_i .

Në përfundim të këtij procesi të dhënat ndodhen brenda segmentit [0,1].

Për realizimin e eksperimenteve janë përdorur 300 rekorde të objekteve të pasurive të paluajtshme (apartamente) të marra gjatë periudhës nga 1 qershor 2019, deri më 1 nëntor 2019, për qytetin e Tiranës. Në tabelën Nr 2 (Rekorde të të dhënave hyrëse të normalizuara) gjenden disa prej rekordeve të normalizuara.

Tabela 2. Rekorde të të dhënave hyrëse të normalizuara

Nr	Nr. i kateve	Nr. i dhomave	Siperfaqja	Zona	Materiali
1	0.33	0.4	0.37	0.4	0.24
2	0.83	0.93	0.12	0.2	0.15
3	1	0.4	0.12	0.2	0.18
4	0.17	0.2	0.25	0.6	0.31
5	0.67	0.67	0.12	0.2	0.15
6	0.5	0.27	1	0.2	0.14
7	0.33	0.33	0.5	0.6	0.27
8	0.17	0.13	0.25	0.6	0.32
9	0.67	0.07	0.62	0.4	0.23
10	0.33	0.2	0.12	0.8	0.95

Rezultatet e eksperimenteve

Janë realizuar dy lloje eksperimentesh duke përdorur parametra të ndryshëm, si: numri i epokave, numri i neuroneve në shtresat e fshehta. Fillimisht,

modeli është mësuar me setin e trajnimit, i cili përbëhet nga të gjitha rekordet dhe po i njëjti set të dhënash është përdorur për testim. Eksperimentet janë realizuar duke përdorur tipologji të ndryshme të arkitekturës së rrjetit neural (Zoff & Le, 2016). Është ndërtuar një rrjet neural me katër shtresa, përkatësisht: shtresa hyrëse, dy shtresa të fshehura dhe shtresa dalëse.

Në figurën Nr 2 (Arkitektura e rrjetit neural) shfaqet një nga arkitekturat e përdorura, numri i neuroneve në shtresën hyrëse, numri i neuroneve tek shtresa e fshehtë dhe neuroni dalës.

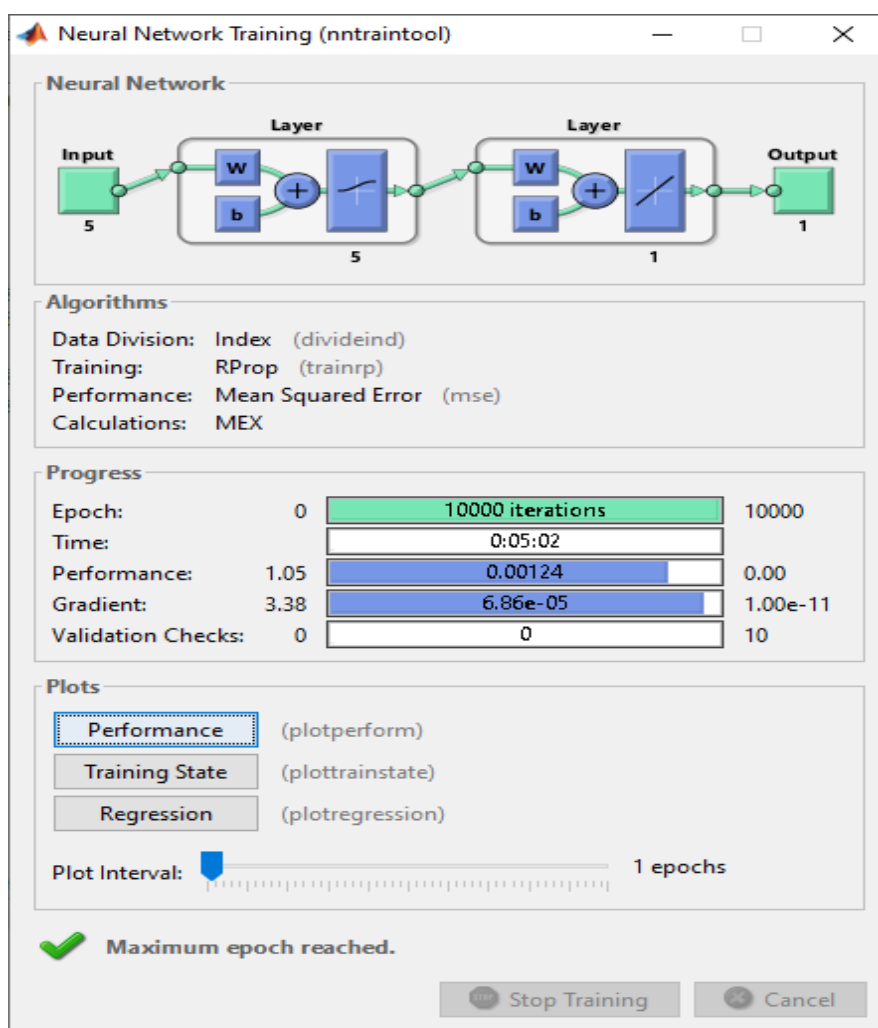


Figura 2. Arkitektura e rrjetit neural

Për zhvillimin e eksperimenteve është përdorur paketa Matlab 8.5 R2015a (Matlab & Simulink). Në eksperimentin e parë, të dhënat e përdorura për trajnim janë përdorur edhe për të testuar modelin. Në eksperimentin e dytë, seti i trajnimit përmban 299 rekordet e para dhe rekordi i fundit u përdor për

testim, pra, testimi u realizua me të dhëna të papërdorura më parë për trajnim. Trajnimi dhe testimi janë përsëritur 299 herë, ku rekordi i patrajnuar përdoret për testim.

Për vlerësimin e performancës së modelit është konsideruar gabimi mesatar kuadratik (Freedman, 2009), i cili përlogarit me formulën e mëposhtme:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Ku n – numri i parashikimeve, Y - vektori i vlerave të vëzhguara të variablit që parashikohet me \hat{Y} - vlera e parashikuar.

Rezultatet, sipas numrit të epokave dhe numrit të neuroneve në shtresën e fshehur, për secilin prej eksperimenteve, paraqiten në tabelën Nr 3 (Rezultatet e eksperimenteve):

Tabela 3. Rezultatet e eksperimenteve

Arkitektura	MSE Eksperiment 1				MSE Eksperiment 2			
	Numri i epokave				Numri i epokave			
Nr. i neuroneve në shtresat e fshehura	5	100	1000	10000	5	100	1000	10000
5	0.1218	0.1171	0.1297	0.1381	0.1887	0.1365	0.1524	0.1606
10	0.1713	0.1375	0.1385	0.1387	0.2424	0.1398	0.1472	0.1432
20	0.2124	0.1384	0.1387	0.1395	0.3692	0.1447	0.1480	0.1497

Gabimi mesatar kuadratik rezulton më i ulët në eksperimentin e parë, ku të dhënat e përdorura për trajnim janë të njëjta me të dhënat e përdorura për testim, si dhe për rrjetin me 5 neurone në shtresat e fshehura dhe për 100 epoka. Në këtë rast, trajnimi dhe testimi përfundoi për 2 minuta, në një makinë me procesor Intel Core i5-8400, x64, me kujtesë fizike 8 GB dhe me sistem operimi Windows 10 Pro.

Rezultatet eksperimentale tregojnë se 100 - 1000 epoka janë të mjaftueshme për trajnim efektiv të rrjetit me 5-10 neurone në shtresën e fshehur. Vihet re një rritje e lehtë e gabimit me rritjen e numrit të neuroneve dhe numrit të epokave, e cila shpjegohet me vëllimin e pamjaftueshëm të të dhënave të trajnimit, ku modeli fillon të memorizojë të dhënat e trajnimit në vend që të mësojë të përgjithësojë prej tyre (Rogers & McLelland, 2004).

Konkluzione

Rrjetet neurale janë të përshtatshme për zgjidhjen shumë problemeve, nëpërmjet të mësuarit nga shembujt, pa shtuar asnjë informacion tjetër i cili edhe nëse është, mund të mos jetë i besueshëm e mund të shkaktojë më tepër

konfuzion sesa parashikim të saktë. Rrjetet neurale artificiale janë të afta të përgjithësojnë (Siddhaling & Prema, 2011) dhe shmangin zhurmat.

Rrjetet neurale artificiale janë përdorur gjërësisht për zgjidhjen e problemeve të ndryshme, sidomos kur nuk zotërohet përshkrim i detajuar i informacionit që përdoret për të mësuar modelin.

Nga eksperimentet rezultoi se gabimi është më i ulët, rreth 10-15%, nëse seti i të dhënave të përdorura për testim është i njëjtë me setin e të dhënave të përdorura për trajnim, ndërsa gabimi në rastin e eksperimentit të dytë (kur seti i të dhënave të përdorura për testim është i ndryshëm nga të dhënat e përdorura për trajnim) është rreth 5% më shumë. Nisur nga këto rezultate, rrjetet neurale mund të shërbejnë si mekanizma të efektshme për zgjidhjen e problemeve të vlerësimit dhe parashikimit me saktësi të lartë.

Literatura

David H. Hubel, Torsten N. Wiesel (2005): Brain and visual perception: the story of a 25-year collaboration. Oxford University Press US. p. 106

Nielsen M. (2015): Neural Networks and Deep Learning. Determination Press

Matlab & Simulink: www.mathworks.com-Introduction to Dynamic Neural Networks - MATLAB & Simulink. (Retrieved 2017-06-15)

Ke-Lin Du, M. N. S. Swamy (2019): Neural Networks and Statistical Learning Second Edition. Springer Nature

Jouke Annema (2012): Feed-Forward Neural Networks: Vector Decomposition Analysis, Modelling and Analog Implementation. Springer US

Jinkun Liu (2013): Radial Basis Function (RBF) Neural Network Control for Mechanical Systems. Springer-Verlag Berlin Heidelberg

Umberto Michelucci (2019): Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection. Apress

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2013): An Introduction to Statistical Learning: with Applications in R. Springer.

Zoph Barret, Le V. Quoc (2016): Neural Architecture Search with Reinforcement Learning. IEEE Access PP(99):1-1

David A. Freedman (2009): Statistical Models: Theory and Practice. Cambridge University Press

Rogers, T., McLelland, J. L. (2004): Semantic Cognition: A Parallel Distributed Processing Approach. Cambridge, MA: MIT Press

Siddhaling U., Prema K. V. (2011): Generalization Capability of Artificial Neural Network Incorporated with Pruning Method. International Conference on Advanced Computing, Networking and Security. ADCONS 2011: Advanced Computing, Networking and Security, p. 171-178

Hardt M., Recht B., Singer. Y. (2016): Train faster, generalize better: Stability of stochastic gradient descent. Proceedings of the 33rd International Conference on Machine Learning, in PMLR 48:1225-1234