

RĪGAS TEHNISKĀ UNIVERSITĀTE
Elektronikas un telekomunikāciju fakultāte
Telekomunikāciju institūta

Gundega RUTKA

Telekomunikāciju un datoru tīklu doktora programmas doktorante

**NEIRONU TĪKLU MODEĻU LIETOŠANA INTERNETA
DATPLŪSMAS PROGNOZĒŠANĀ**

Promocijas darba kopsavilkums

Zinātniskais vadītājs
Dr.sc.ing., profesors
G.LAUKS

Rīga 2008

UDK 338.24 (075.8)
E1 710

Rutka G. Neironu tīklu modeļu lietošana interneta datplūsmas prognozēšanā. Promocijas darba kopsavilkums.-R.:RTU,2008.-24 lpp.

Iespiests saskaņā ar RTU ETF promocijas padomes „RTU P-08” 2008.gada 12.marta lēmumu, protokols Nr.1

Šis darbs izstrādāts ar Eiropas Sociālā fonda atbalstu Nacionālās programmas „Atbalsts doktorantūras programmu īstenošanai un pēcdoktorantūras pētījumiem” projekta „Atbalsts RTU doktorantūras attīstībai” ietvaros”.

Promocijas darba galvenie rezultāti ir izskaitļoti Grid tīklā, izmantojot RTU ETF Grid klāstera skaitļošanas jaudas.

ISBN

PROMOCIJAS DARBS
IZVIRZĪTS RĪGAS TEHNISKAJĀ UNIVERSITĀTĒ
INŽENIERZINĀTŅU (telekomunikāciju)
DOKTORA GRĀDA IEGŪŠANAI

Promocijas darbs inženierzinātņu (telekomunikāciju) doktora grāda iegūšanai tiek publiski aizstāvēts 2008.gada 5.jūnijā Rīgas Tehniskās universitātes Elektronikas un telekomunikāciju fakultātē, Āzenes ielā 12, 210. auditorijā.

OFICIĀLIE OPONENTI:

Profesors, dr.sc.comp. Rihards Balodis- Bolūžs

Docents, dr.sc.comp. Modris Greitāns

Profesors, dr.sc.ing. Ernests Pētersons

APSTIPRINĀJUMS

Es apstiprinu, ka esmu izstrādājusi doto promocijas darbu, kas iesniegts izskatīšanai Rīgas Tehniskajā universitātē inženierzinātņu doktora grāda iegūšanai. Promocijas darbs nav iesniegts nevienā citā universitātē zinātniskā grāda iegūšanai.

Gundega Rutka (paraksts)

Datums:.....

Promocijas darbs ir uzrakstīts latviešu valodā, satur ievadu, 6 nodaļas, secinājumus, 4 pielikumus, 62 attēlus un 55 tabulas pamattekstā, kopā 136 lapaspuses. Bibliogrāfiskajā sarakstā ir 109 nosaukumi.

Tēmas aktualitāte

Šodienas telekomunikāciju tīkli ir ļoti sarežģīti – pēc pielietojamām tehnoloģijām, tīkla arhitektūras, topoloģijas utt. Tas norāda, ka pārraidot informāciju tīklā, kurš sastāv no vairākiem tīkla līmeņiem, ir sarežģīti pētīt datplūsmas¹ statistiku un līdz ar to analizēt un simulēt tīkla datplūsmu. Tāpat tīklā tiek ieslēgtas liela apjoma aplikācijas, pastāv dažādas iespējamības tīkla savietojamībai, arhitektūrai un iekārtām, kas ieslēgtas tīklā, līdz ar to datplūsmas raksturs kļūst grūti aprakstāms. Līdz ar to var teikt, ka vienmēr pastāvēs kādas modifikācijas, dēļ kā datplūsmas simulācija kļūst sarežģīta [6].

Viens no svarīgākajiem elementiem uz mērījumiem balstītā tīkla vadībā ir paredzēt datplūsmu nākošajā laika intervālā, kad tiks veikta “on line” - tiešsaistes - režīma trafika uzskaitē. Galvenais uzdevums, ņemot vērā vēsturisko statistiku, ir paredzēt datplūsmu un tās izmaiņas pēc iespējas precīzāk. Datplūsmas prognozei ir nepieciešami precīzi datplūsmas modeļi, kuri var veikt statistisku uzskaiti esošajai datplūsmai. Jo neprecīzāks modelis, jo nepilnīgāka un kļūdaināka tīkla datplūsmas prognoze. Pastāv dažādi tīkla datplūsmas modeļi. Ir nepieciešams izveidot prognozētāju, kuram ir šādi kritēriji:

- 1) precizitāte- vissvarīgākais kritērijs, jo balstoties uz to, varēs iegūt nākotnes vērtību prognozes modeli;
- 2) vienkāršība- lai iegūtu reālā laika prognozētāju ir nepieciešama zināma vienkāršība lietošanā.

Pakešu komutācijas tīklu datplūsma, piemēram, kā interneta datplūsma, mūsdienās tiek raksturota citādāk par tradicionālo telefonu tīklu datplūsmu. Vairāku autoru darbos [11], [15], [19], [26], [16] tiek norādīts, ka liela ātruma tīklos tiek novērota pašlīdzība, ko nevar novērot ar klasiskajiem datplūsmas modeļiem. Respektīvi, kā galvenais interneta datplūsmas raksturojošais elements parādās fraktālis, t.i., ilgtermiņa atkarība², īstermiņa atkarība³, autokorelācija, pašlīdzība⁴ utt. Šie lielumi raksturo tīkla veiktspēju un tīkla sarežģīto uzbūvi. Daudzi autori ir veltījuši pastiprinātu uzmanību pašlīdzīgas datplūsmas analīzei un novērtēšanai [28], [2], [18]. Jāpiebilst, ka joprojām nav vienota viedokļa par pielietoto metožu precizitāti.

Pētījumu mērķis – Izstrādāt Neironu tīklu modeļus interneta datplūsmas prognozēšanai tīkla aplikācijas līmenī.

Lai sasniegtu promocijas darbā nosprausto mērķi, ir jāveic šādi uzdevumi:

1. Jāizanalizē pašlīdzības efekta novērtēšanas metodes, lai konstatētu, kura metode ir piemērojama interneta datplūsmas analīzei tīkla aplikācijas līmenī.
2. Jāsagatavo interneta datplūsmas trases ar dažādiem nolases soļiem un dažādos laika intervālos, jānosaka to pašlīdzības pakāpe (Hersta koeficients), autokorelācijas funkcija, lai prognozēšanas uzdevums tiktu veikts ar dažādām trasēm (īstermiņa, ilgtermiņa process) un varētu novērtēt rezultātus dažādiem procesiem.
3. Jāapkopo informācija par neironu tīklu modeļiem, kurus ir mērķtiecīgi izmantot interneta datplūsmas prognozēšanai un jāpielāgo tie prognozēšanas uzdevumam.
4. Jāizpēta un jānovērtē neironu tīklu modeļu prognozēšanas kļūdas, to precizitāte.
5. Jāizstrādā rekomendācijas pašlīdzīgas datplūsmas prognozēšanai.
6. Jāizstrādā procedūra pašlīdzīgas datplūsmas prognozēšanai.

¹ pēc Latvijas ZA Terminoloģijas komisijas tulkojuma angļu val.vārdam “traffic” - datu, ziņojumu vai citas digitālas informācijas plūsma sakaru kanālā vai tīklā.

² Angļu valodā – “Long range dependence- LRD”.

³ Angļu valodā – “Short-range dependence- SRD”.

⁴ Angļu valodā – “Self- similarity”.

Pielietotās pētīšanas metodes. Promocijas darba izstrādes laikā tika pielietotas šādas pētīšanas metodes, kas tika realizētas ar Matlab vers.6.5, Statistika vers.5.5. un MSExcel programmatūru nodrošinājumu:

1. Laika- dispersijas metode pašlīdzīgas datplūsmas novērtēšanai.
2. R/S statistiskā metode pašlīdzīgas datplūsmas novērtēšanai.
3. Neironu tīklu modeļi apmācības un prognozēšanas uzdevumam.
4. Matemātiskās statistikas metodes prognozēšanas kļūdas ticamības intervāla atrašanai.

Zinātniskais jauninājums. Promocijas darbā tika iegūti šādi jauni zinātniskie rezultāti:

Tika izveidoti neironu tīklu modeļi interneta datplūsmas prognozēšanai. Ņemot vērā datplūsmas pašlīdzīgo raksturu, prognozēšanai tika izmantoti lineārie, vienvirziena un radiālās bāzes funkcijas tīkli.

Promocijas darbā tika apkopoti šādi jauni praktiskie rezultāti:

Tika izpētīta interneta datplūsmas prognozēšanas kļūda, izmantojot dažādus neironu tīklu modeļus. Vislabākie datplūsmas prognozēšanas modeļi ir lineārie tīkli. Vienvirziena tīklos prognozēšanas kļūda ir lielāka nekā lineārajos tīklos, bet RBF tīklu skaitļošana prasa lielus datora atmiņu resursus un lielu skaitļošanas laiku. Tika izstrādātas rekomendācijas interneta datplūsmas prognozēšanas modeļa izveidošanai. Balstoties uz šīm prasībām, tika izstrādātas vadlīnijas interneta datplūsmas prognozēšanai ar neironu tīklu modeļiem.

Darba praktiskā vērtība ir rezultāti, kas tika iegūti promocijas darba izstrādes laikā. Šos rezultātus rekomēdēju elektronisko sakaru tīklu attīstības plānošanas sistēmā, lai veiktu īslaicīgu/ ilglaicīgu trafika pieauguma novērtējumu, prognozētu aparatūras iegādi un kanālu paplašināšanu. Iegūtie rezultāti ir noderīgi operatoriem, kas specializējas mājas lapu izvietojumā un uzturēšanā Interneta vidē (hostinga pakapojumi). Jo vairāk apmeklētāju konkrētām mājas lapām, jo lielāka slodze uz serveri, kas var radīt būtiskus servera darbības traucējumus vai arī pārtraukumu vispār. Hostinga operatoriem ir svarīgi nodrošināt piekļuvi pakalpojumiem, lai lietotājs būtu apmierināts, jo piekļuve konkrētām globālā tīmekļa mājas lapām veido būtisku e-komercijas daļu. Prognozēšanas rezultāts dod iespēju savlaicīgi plānot elektronisko sakaru tīklu paplašināšanu.

Aprobācija

Promocijas darba galvenie rezultāti tika nolasīti uz apspriesti:

1. The 9th Biennial Baltic Electronic Conference, Tallina, Igaunija, **2004.** gada 3.-6.oktobrī
2. RTU 46.starptautiskā zinātniskā konference, Rīga, Latvija, **2005.** gada 13.-14.oktobrī
3. 10th international conference „Electronics’ 2006”, Viļņa, Lietuva, **2006.** gada 23.-25.maijā
4. 11th international conference „Electronics’ 2007”, Kauņa, Lietuva, **2007.** gada 15.-17.maijā

Kopumā tika publicēti četri un pieņemti publicēšanai divi zinātniskie raksti dažādos zinātniskos izdevumos:

1. G.Rutka, „Traffic Prediction In Mobile Networks Using Neural Network Applications”, Proceedings of the 9th Biennial Baltic Electronic Conference, pp.303-306, **2004**
2. G.Rutka, „Some Experimental Results Of Internet Traffic Prediction Using Neural Network”, RTU zinātnisko rakstu krājums, Telekomunikācijas un Elektronika, Sērija7, Sējums 5, pp.6-9, **2005**
3. G.Rutka, „Neural Network Models For Internet Traffic Prediction”, Electronics and Electrical Engineering - Kaunas, Nr.4(68), pp.55-58, **2006**
4. G.Rutka, G.Lauks „Study on Internet Traffic Prediction models”, Electronics and Electrical Engineering - Kaunas, Nr.6(78), pp.47-50, **2007**

5. G. Rutka, „Network Traffic Prediction Using Neural Networks”, Electronics and Electrical Engineering - Kaunas, Nr. **2008** (pieņemta publicēšanai)
6. G. Rutka „Prediction Accuracy of Neural Network Models”, Electronics and Electrical Engineering - Kaunas, **2008** (pieņemta publicēšanai)

Darba struktūra

Promocijas darbs sastāv no ievada, 6 nodaļām, secinājumiem, bibliogrāfijas un 4 pielikumiem.

Ievadā ir pamatota veikto pētījumu aktualitāte, formulēts darba mērķis un uzdevumi, uzskaitītas promocijas darba izstrādē lietotās zinātniskās metodes, aprakstīta pētījumu zinātniskā novitāte un iegūto rezultātu praktiskā nozīmība, kā arī ir raksturota darba aprobācija.

Promocijas darba 1.nodaļā ir apkopota informācija par interneta datplūsmas raksturojumu mūsdienu telekomunikāciju tīklos. Īpaša uzmanība tiek veltīta pašlīdzības noteikšanas metodēm, to precizitātei. Tiek apskatīta autokorelācijas nozīme prognozēšanā, kā arī interneta datplūsmas raksturīgākie laika intervāli.

Darba 2.nodaļa tiek veltīta pētījumos izmantoto datplūsmas datu apskatei. Tiek novērtēta datplūsmas datu pašlīdzība pēc dispersijas- laika metodes un R/S statistiskās metodes. Iegūtie rezultāti tiek salīdzināti.

Darba 3.nodaļa runā par pētījumos izmantoto datplūsmas datu autokorelācijas izpēti. Balstoties uz darba 3.nodaļā iegūtajiem rezultātiem, darba 4.nodaļā tiek formulēta prognozēšanas problēma. Tiek apskatīti ARIMA/ FARIMA un neironu tīklu prognozēšanas modeļi. Šajā nodaļā sīkāk tiek apskatītas prognozēšanas kļūdas novērtēšanas metodes.

Darba 5.nodaļa tiek veltīta prognozēšanai ar neironu tīkla modeļiem. Šajā nodaļā ir apkopota informācija par izmantotajiem prognozēšanas principiem, iegūtajiem rezultātiem. Balstoties uz darba 5.nodaļā iegūtajiem rezultātiem, darba 6.nodaļa piedāvā neironu tīklu modeļu izvēles algoritmu interneta datplūsmas prognozēšanai.

Promocijas darba galvenie secinājumi ir apkopoti darba noslēguma nodaļā.

1. Interneta tīkla datplūsmas raksturojums

Mūsdienās sabiedrības ikdienā informācija ienāk aizvien vairāk un vairāk. Pakešu komutācijas tīklos parādās tādi jēdzieni kā paketēšana, tīkla slāņi un protokoli. Tas nozīmē, ka parādās invarianti⁵, kas vairāk tiek saistīti ar tīkla slāņu raksturojumu, bet kas pēc savas idejas ir tālu no datplūsmas statistikas, respektīvi, ir daudz sarežģītāk analizēt un simulēt tīkla datplūsmu. Mūsdienu interneta tīkls ir savstarpēji savienots ar daudzām un dažādām aplikācijām (ne tikai ar balss pārraides iekārtām), kur katrai no aplikācijām (iekārtām) ir savs raksturojums. Pie kam, šo aplikāciju skaits var nemitīgi mainīties. Līdz ar to kļūst sarežģītāk analizēt un raksturot tīkla datplūsmu. Tāpat dažādas atšķirības parādās tīkla savietošanai, atšķirīga tīkla arhitektūra, dažādas datplūsmas plūsmas, kas kopsummā dod rezultātu- vienu datplūsmas statistiku, kuru ir sarežģīti analizēt, tāpēc, ka šī statistika ir veidojusies no dažādiem apstākļiem un likumsakarībām [6]. Un viena no galvenajām problēmām, kas parādās pakešu komutācijas tīklos, ir datplūsmas „uzliesmojošais” raksturs⁶ dažādos laika šķērsgriezumos [7]. Izpētot šo problēmu sīkāk Bellcore pētnieki nonāca pie secinājuma, ka

⁵ Pēc LZ Terminaloģijas komisijas tulkojuma angļu val. vārdam „invariant” - matemātisks vai programmisks objekts, kas paliek nemainīgs kādu pārveidojumu rezultātā.

⁶ Angļu valodā - “burstiness”.

šāds tīkla datplūsmas „uzliesmojošais” raksturs ir novērojams dažādos laika mērogos, pie kam, lielākos laika mērogos šī parādība atkārtojas. Sākotnēji plaši eksperimenti tika veikti lokālajā tīklā⁷ [6].

Promocijas darba izstrādē tiek veikti pētījumi ar interneta datplūsmu, kas ievāka aplikācijas līmenī (diskrēti gadījuma lielumi). Līdz šim pašlīdzības efektu vairāk pētīja lokālajā tīklā [11], [25] un teritoriālajā tīklā⁸[18], Interneta tīklos kopumā [1]. Līdzšinējie pētījumi skaidri neidentificē pašlīdzības efektu interneta datplūsmai konkrētā tīkla līmenī.

Ir novēroti liela daudzuma datplūsmas avoti, kas darbojas pēc pašlīdzības principa lielos laika mērogos, piemēram, interneta vidē [18], [15]. Vislabāk pašlīdzību apraksta ar tā saucamajiem datplūsmas sakopojumiem⁹.

Pieņemam, ka $X = \{X_t; t=0, 1, 2, \dots\}$ ir stacionārs stohastisks process ar vidējo lielumu μ , dispersiju σ^2 un autokorelācijas funkciju r_k , kur $k=0, 1, 2, \dots$, tad autokorelācijas funkcija no X ir šāda:

$$r_k \sim k^{-(2-2H)}L(k) \text{ jeb } r_k \sim k^{-\beta}L(k) \text{ pie } k \rightarrow \infty, \quad (1.1)$$

kur H tiek saukts par Hersta parametru. $L(k)$ apzīmē funkciju, kas lēnām mainās uz bezgalību:

$$\lim_{z \rightarrow \infty} \frac{L(zt)}{L(t)} = 1, \text{ pie visiem } z > 0. \quad (1.2)$$

Tiek pieņemts, ka $L(k)$ ir konstanta vērtība 1. Hersta parametrs H ir diapazonā $0.5 < H < 1$ un raksturo procesa pašlīdzības pakāpi un ilgtermiņa atkarību. Pašlīdzības un ilgtermiņa atkarības pakāpe palielinās, ja $H \rightarrow 1$. Tas nozīmē, ka, ja procesam $H=0.5$, tad procesam ir vāji izteikta pašlīdzība, bet, ja H ir tuvu vērtībai 1, tad procesam piemīt izteikta pašlīdzība. Tas nozīmē, ka procesam ir ilgtermiņa atkarība – procesam pagātnē ir bijusi tendence palielināties (vai samazināties), tātad arī nākotnē pastāv augsta varbūtība, ka procesam būs tendence palielināties (vai samazināties) [19], [20], [9], [21].

Katram $m=1, 2, 3, \dots$, pieņemam, ka $X^{(m)} = \{X_k^{(m)}; k=1, 2, 3, \dots\}$ apzīmē jaunas laika sērijas X , kuras nedublējas ar blokiem, kuru izmēri ir m , tas ir, katram $m=1, 2, 3, \dots$, $X_k^{(m)}$ ir uzdots [27]:

$$X_k^{(m)} = \frac{X_{km-m+1} + \dots + X_{km}}{m}, \text{ kur } k=1, 2, 3, \dots \quad (1.3)$$

Vienkāršojot formulu (1.3), iegūstam:

$$X_k^{(m)} = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \text{ kur } k=1, 2, 3, \dots \quad (1.4)$$

Laika sērijas $X_k^{(m)}$ uzskata par izteikti¹⁰ jeb asimptotiski pašlīdzīgām pie visiem $m > 0$ ar parametru β ($0 < \beta < 1$), ja dispersija ir:

$$\text{VAR}(X^{(m)}) = \text{VAR}(X)/m^\beta \quad (1.5)$$

un autokorelācijas funkcija ir:

$$r_k^{(m)} = r_k \text{ pie } m \rightarrow \infty \quad (1.6)$$

Hersta koeficientu aprēķina šādi:

$$H = 1 - \beta/2 \quad (1.7)$$

⁷ Angļu valodā – “Local Area Network- LAN”.

⁸ Angļu valodā – “Wide Area Network – WAN”.

⁹ Angļu valodā – “cluster”.

¹⁰ Angļu valodā – “exactly”.

Pašlīdzīgu datplūsmu var analizēt pēc vairākām metodēm: 1) dispersijas laika metode¹¹, 2) izvērstā mēroga metode (R/S statistiskā metode)¹², 3) periodogramma¹³, 4) Vaita novērtējuma metode¹⁴, 5) apgabalu analīze¹⁵ un 6) teksteļu grafiskā metode¹⁶ u.c. [6]. Promocijas darbā tiks analizētas Hersta koeficienta noteikšanas metodes ar dispersijas- laika metodi un izvērstā mēroga jeb R/S statistisko metodi. Šīs divas metodes izvēlētas, ņemot vērā veicamos aprēķinus, metodes vienkāršību un precizitāti. Vairums Hersta noteikšanas metodes balstās un grafiskā atainojuma analīzi, kas ir ērtākais variants gadījumos, ja analizējamā datplūsma satur dažādu informāciju vienkopus, piemēram, sesiju skaitu, sesijas ilgumu, TCP savienojumu skaitu, starpintervālu laiku starp paketēm utt. Tā kā eksperimenti tiek veikti ar datplūsmu, kuras vienīgā mērvienība ir pieprasījumu skaits laika vienībā, tad tiek pieņemts, ka iepriekš minētās metodes ir vispiemērojamākās.

1.2.1. Dispersijas-laika metode

Daudzi autori, kas ir veltījuši pastiprinātu interesi pašlīdzīgas datplūsmas izpētē, līdz pat šai dienai nav noākuši pie vienota viedokļa par pašlīdzības efekta novērtēšanas metožu precizitāti [28], [2], [18]. Balstoties uz autoru veiktajiem eksperimentu rezultātiem, vislabāko pašlīdzīgas datplūsmas novērtējumu dod dispersijas- laika metode, ko arī izmantosim savos eksperimentos un mērījumos.

Pašlīdzīga procesa dispersija tiek definēta formulā (1.5). Savukārt Hersta koeficients tiek aprēķināts pēc formulas (1.7). Formulu (1.7) var pārrakstīt šādā formā:

$$\log\{\text{VAR}(X^{(m)})\} \sim \log\{\text{VAR}(X)\} - \beta \log\{m\} \quad (1.8)$$

Ja $\text{VAR}(X)$ un m tiek attēloti logaritmiskajā mērogā, tad izmantojot 2.pakāpes lineārās regresijas līkni tiek iegūta līkne $-\beta$, kuras vērtību viegli var noteikt.

1.2.2. R/S statistiskā metode

Izvērstā mēroga metode [17], [2] ir balstīta uz statistikas analīzi izvērstā mērogā R/S. No sākotnējās procesa X_t nolašu secības $\{X_k\}_{k=1}^n$ tiek definēta starpību summu W_k :

$$W_k = \sum_{i=1}^k X_i - k\bar{X}_n, \quad (1.9)$$

kur \bar{X}_n ir aritmētiski vidējā vērtība procesa X_t n nolasēm:

$$\bar{X}_n = \frac{1}{n} \sum_{t=1}^k X_t, \quad (1.10)$$

Ņemot vērā procesa $X(t)$ n nolašu dispersiju $S(n)$ un starpību summu W_k , var aprakstīt izvērstā mēroga statistiku:

¹¹ Angļu valodā – “variance-time plot”.

¹² Angļu valodā – “rescaled range (R/S)”.

¹³ Angļu valodā – “periodogram”.

¹⁴ Angļu valodā – “Whittle estimator”.

¹⁵ Angļu valodā – “wavelet analysis”.

¹⁶ Angļu valodā – “texture plots”.

$$\frac{R(n)}{S(n)} = \frac{\max(0, W_1, \dots, W_n) - \min(0, W_1, \dots, W_n)}{S(n)} \quad (1.11)$$

Ņemot vērā formulu (1.11.) var uzrakstīt tās sakarību ar Hersta parametra noteikšanu:

$$\frac{R(n)}{S(n)} \sim \frac{n^{\alpha H}}{2} \quad (1.12)$$

Logaritmējot formulu (1.12) abās izteiksmes pusēs tiek iegūts:

$$\log_{10} \frac{R(n)}{S(n)} = \alpha H \log_{10} n - \alpha H \log_{10} 2 \quad (1.13)$$

Logaritmiskajā mērogā tiek iegūta līkne ar slīpumu αH un sākuma koordinātu $-\alpha H \log_{10} 2$.

1.3. Autokorelācija

Eksistē daudz prognozēšanas metožu, kurās izmanto autokorelācijas analīzi (korelācijas koeficientus). Autokorelācijas funkcijai ir būtiska nozīme, raksturojot datplūsmu [12], [13]. Ja autokorelācijas funkcija krītas lēnām, tiecoties uz 0, tad noteikti var runāt par pašlīdzīgu procesu. Autokorelācijas funkciju diskrētiem gadījuma lielumiem aprēķina šādi:

$$r(k) = \frac{\frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n)}{\frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2} \quad (1.14)$$

kur

$$\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t \quad (1.15)$$

Autokorelācijas pētīšana ir būtiska prognozēšanas modeļa izvēlē. Ja datplūsmai autokorelācija ir 0, tad nekāda prognozēšanas modeļu izvēle nevar būt. Acīmredzot vislabākais prognozētājs šajā gadījumā būs vidējā nolases vērtība. Sīkāk autokorelācijas līknes skatīt 3.nodaļā.

Šajā nodaļā tika konstatēts:

- Līdzšinējos pētījumos nav konkrētas rekomendācijas pašlīdzības efekta analīzei interneta datplūsmai konkrētā tīkla slānī. Disertācijas darbs tiks veltīts pašlīdzības efekta analīzei aplikācijas slānī (pēc OSI).
- Datplūsmas pašlīdzības efekta novērtēšanā nav vienota viedokļa par pielietoto metožu precizitāti. Disertācijas darbā tiks izanalizēta dispersijas- laika metode un izvērstā mēroga jeb R/S statistiskā metode.
- Lietotāju uzvedība, datu ģenerēšana un pieprasījumi konkrētām mājas lapām, vai ielādes failiem tiek raksturota sākot no dažām sekundēm līdz pat vairākām minūtēm un stundām. Disertācijas darbā ir jāanalizē dažādas nolašu trases ar dažādiem laika soļiem, piemēram, ar laika soli 1 sek un 1h.

- Autokorelācijai ir būtiska loma tīkla datplūsmas raksturošanā. Tās analīze ir nozīmīga prognozēšanas modeļa izvēlē. Disertācijas darbā ir jānosaka autokorelācijas funkcijas vērtība (jābūt lielākai par 0, lai varētu veikt prognozēšanu). Vizuāli jānovērtē autokorelācijas līknekāds ir raksturīgais periodiskums, ja tāds eksistē.
- Interneta datplūsmas analīzes pamatā ir datizrace. Disertācijas darbā apskatāmajām datu nolāsēm ir jāatrod likumsakarības, pēc kurām var veidot prognozēšanas modeļus. Likumsakarību atrašanai tiks izmantoti neironu tīkli.

2. Pētījumos izmantoto datplūsmas datu apskats un pašlīdzības pakāpes noteikšana

Līdz šim dažādi autori ([4], [21], [27]) pašlīdzības efektu novērtē imitētajām datu nolāsēm. Šajā promocijas darbā Hersta koeficienta noteikšanai tiek izmantotas dažādas nolases – reālas statistikas datu nolases un imitētās datu nolases, kuras īsi ir aprakstītas 2.1.tabulā.

2.1.tabula

Datu nolašu raksturojums

Nr. p.k.	Nosaukums	Nolašu skaits	Solis/ laika intervāls	Kopējais nolašu periods
1.	Freestats trase ¹⁷	168	1h	7 dienas
2.	Freestats trase	1176	1h	49 dienas
3.	Freestats trase	1272	1h	53 dienas
4.	Freestats trase	1986	1h	82 dienas
5.	Freestats trase	8760	1h	365 dienas
6.	Fotoblog trase	172800	1s	2 dienas
7.	MMPP Freestats trase (49d)	1176	1s	imitēta datplūsma
8.	MMPP Freestats trase (82d)	1986	1s	imitēta datplūsma
9.	MMPP Freestats trase (365d)	8760	1s	imitēta datplūsma
10.	MMPP Fotoblog trase	172800	1s	imitēta datplūsma

Ņemot vērā vairāku autoru darbus [14], [22], pašlīdzības datplūsmas imitācijai par pamatu esam ņēmuši MMPP. MMPP ir plaši izmantots B-ISDN avotu raksturošanai, piemēram, balss un video pārraidei, kā arī interneta datplūsmas „uzliesmojošās” dabas raksturošanai [5], [23], [29], [42], [10], [24].

Apstrādājot simulētas datplūsmas datus, Hersta parametra novērtēšanas rezultāts dažādām metodēm ir līdzīgs. Vairums dažādu publikāciju autori, pētot pašlīdzīgu datplūsmu, tās novērtēšanu, ir darbojušies tikai ar teorētiski simulētu datplūsmu [15], [27].

2.2.tabulā ir apkopoti eksperimentos iegūtie rezultāti Hersta parametra noteikšanai ar dispersijas- laika un R/S statistisko metodi.

¹⁷ Pēc LZ Terminoloģijas komisijas tulkojuma angļu val. vārdam „trace” - Programmas izpildāmo instrukciju un mainīgo vērtību, kā arī cita veida informācijas, kas saistīta ar programmas izpildi, pieraksts.

Hersta koeficientu salīdzinājums

Nr.p.k.	Datplūsma	$H_{\text{dispersijas-laika}}$	$H_{R/S}$	Atšķirība*
1.	Freestats trase (7d)	0.57	0.63	10.53%
2.	Freestats trase (49d)	0.64	0.66	3.13%
3.	Freestats trase (53d)	0.64	0.66	3.13%
4.	Freestats trase (82d)	0.71	0.66	7.04%
5.	Freestats trase (365d)	0.75	0.66	12.00%
6.	Fotoblog trase	0.87	0.55	36.78%
7.	MMPP Freestats trase (365d)	0.56	0.63	12.50%
8.	MMPP Fotoblog trase	0.53	0.70	32.08%

* atšķirība tiek aprēķināta šādi:
$$R = \frac{H_{\text{dispersija s-laika}} - H_{R/S}}{H_{\text{dispersija s-laika}}} \cdot 100, \%$$

Salīdzinot veiktos aprēķinus, jāsecina, ka pašlīdzības pakāpes noteikšana pēc dispersijas- laika metodes un R/S statistiskās metodes atšķiras no 3% līdz pat 37%. Lielāka atšķirība parādās datu trasēm, kuras satur lielus nolašu skaitus.

Pēc dispersijas- laika metodes Hersta koeficients palielinās, palielinoties analizējamo nolašu skaitam. Kopumā, apskatāmās datu trases ir pašlīdzīgas, jo Hersta koeficients iekļaujas diapazonā $0.5 < H < 1$. Vislielākais Hersta koeficients tiek sasniegts reālai datu nolasei- Fotoblog trasei. No tā var secināt, ka šī datu trase ir pašlīdzīga ar ilgtermiņa atkarību. Datu trases, kurām Hersta koeficients ir tuvāk vērtībai 0.5 arī tiek uzskatītas par pašlīdzīgām, bet ar īstermiņa atkarību.

Hersta koeficients pēc R/S statistiskās metodes reālām datu nolasēm svārstās no 0.55-0.66. Arī pēc R/S statistiskās metodes Hersta koeficients gan reālām datu nolasēm, gan imitētajām datu nolasēm ir atšķirīgs. Kopumā iegūtie rezultāti pēc R/S statistiskās metodes ļauj secināt, ka analizējamās datu nolasēs ir pašlīdzīgas.

R/S statistiskās metodes aprēķini aizņem lielāku laiku, jo prasa ļoti sīkus aprēķinus. Tāpat ar R/S statistisko metodi ne visos gadījumos ir iespējams veikt aprēķinus pie līdzīgām apkopojuma pakāpēm m^{18} , kādus izmanto dispersijas - laika metode. Lai gan R/S statistiskā metode pieļauj izmantot dažādas un pat vairāk apkopojuma pakāpes nekā dispersijas- laika metode (tikai tādas apkopojuma pakāpes, kurās apskatāmo nolašu skaits ir vesels skaitlis), iegūto gala rezultātu tas būtiski neietekmē.

Analizējot Fotoblog trasi vai MMPP Fotoblog trasi pēc R/S statistiskās metodes, jākonstatē, ka to nav iespējams korekti izdarīt, jo vairums gadījumos, it īpaši pie mazākām apkopojuma pakāpēm (piemēram, no 2 līdz pat 200), tādu starprezultātu kā R un S (skatīt formulu (1.11)) rezultāts ir 0. Pēc R/S statistiskās metodes logaritmiskajā mērogā ir jāuzzīmē R un S dalīšanas rezultāts, kas šajos gadījumos ir 0. Taču logaritms no 0 ir neiespējams, jo paši logaritmi var būt gan pozitīvi (ja skaitļi ir lielāki par 1, vai negatīvi (ja skaitļi ir starp 0 un 1). Līdz ar to pie šīm apkopojuma pakāpēm nevar iegūt korektus rezultātus.

¹⁸ Angļu valodā – “aggregation level”.

Tā kā iegūtie rezultāti liecina par pašlīdzīgu datuplūsmu, tad jāsecina, ka pašlīdzība ir konstatējama datuplūsmai, kas ievākta gan ar soli 1h, gan ar soli 1 sek.

Līdz ar to:

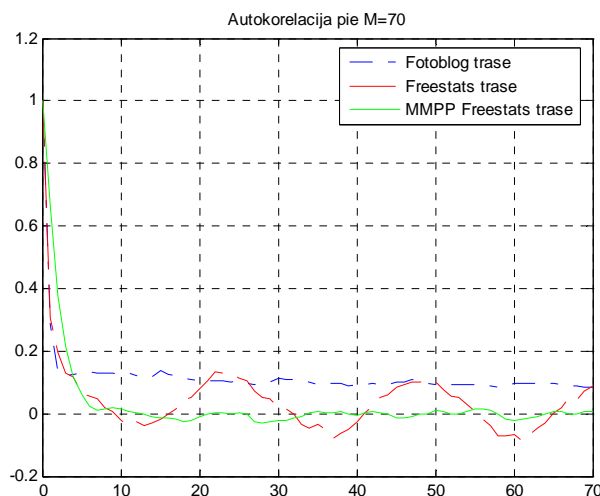
- R/S statistiskā metode uzrāda neprecīzus rezultātus datu trasēm, kas satur secīgas vērtības „0” un datu nolašu vērtībām ir tendence minimāli mainīties (piemēram, Fotoblog trase un MMPP Fotoblog trase). Tas izskaidrojams ar to, ka starprezultātu kā R un S (skatīt formulu (1.11)) rezultāts ir 0, taču logaritmējot šo rezultātu attiecību R/S nav iespējams rezultāts, tā kā logaritms no 0 ir neiespējams.

- dispersijas –laika metodi var pielietot jebkurai datu nolasei. Pašlīdzības koeficienta novērtēšanas aprēķinos ir jāizmanto visas iespējamās apkopojuma pakāpes m, lai gala rezultāts būtu korekts.

Tā kā interneta datuplūsmas statistika (pieprasījumu skaits laika vienībā) var saturēt secīgas vērtības „0”, kā arī datu nolašu vērtības var minimāli mainīties, korektu rezultātu iegūšanai pašlīdzīgas datuplūsmas pakāpes noteikšanā (Hersta parametrs) tiek rekomendēta dispersijas – laika metode. Šī metode ir viegli pielietojama pēc metodes vienkāršības veida, veicamo aprēķinu daudzuma un izpildes ātruma.

3. Pētījumos izmantoto datuplūsmu autokorelācijas izpēte

3.nodaļa satur detalizētu autokorelācijas funkcijas (AKF) līkņu izpēti. Tā kā apskatāmajām datuplūsmām ir visai liels nolašu skaits un kopējā attēlā nav uzskatāmi redzamas atšķirības, tad AKF tiek apskatīta pie mazāka nolašu skaita - skatīt 3.1.attēlu).



3.1.att. Autokorelācija Fotoblog, Freestats un MMPP Freestats trasei pie 70 nolasēm

3.1.attēlā ir redzams, ka pie 70 nolasēm datuplūsmu AKF raksturs ir eksponenciāls. Freestats trases līkne ir lēni rimstoša sinusa līkne. Šajā attēlā nepārprotami var redzēt, ka Freestats trasei autokorelācijas līkne ir cikliska. Novērtējot cikliskumu, iegūst, ka $T=24$ vienības, kas atbilst 24h. Fotoblog trases AKF satur tikai pozitīvas vērtības, savukārt MMPP Freestats trase eksponenciāli krītas, vērtībai svārstoties ap vērtību 0.1.

Izpētot AKF līknes, var sniegt šādus secinājumus:

- MMPP autokorelācijas līkne krītas eksponenciāli. Pēc šīs līknes var secināt, ka process ir pašlīdzīgs ilgtermiņa process.

- Fotoblog trases autokorelācijas līkne ir līdzīga Puasona procesa autokorelācijas līknei, taču tā pieņem gandrīz nemainīgu vērtību, kas svārstās ap 0.1. Pēc šīs līknes var spriest, ka process ir pašlīdzīgs ilgtermiņa process.

- Freestats trases autokorelācijas līkne eksponenciāli krītas, taču pati līkne atgādina sinusa funkcijas līkni. Respektīvi, šī autokorelāciju nepārprotami uzrāda, ka datplūsmai ir ciklisks raksturs. Papētot šo autokorelāciju dažādos laika mērogos, var aprēķināt cikla periodu, kas atbilst 24 nolasēm. Šis process ir pašlīdzīgs īstermiņa process.

No tā jāsecina, ka:

- Izvēloties apmācības un prognozēšanas modeli, īstermiņa un ilgtermiņa procesam apmācības vektors un prognozēšanas vektora garums ir jāizvēlas atbilstoši autokorelācijas periodam.

4. Prognozēšanas problēmas formulējums

Datplūsmas prognozēšanā tiek izvirzīti šādi būtiski aspekti : 1) prognozēšanas metode un 2) prognozēšanas intervāls. Datplūsmas prognozēšanai, lai tā būtu pietiekoši efektīva, ir jābūt vienkāršai. Savukārt prognozes intervāls nosaka, cik tālu nākotnē ir vēlme ieskatīties un ar kādu ticamības pakāpi veiksīm datplūsmas prognozēšanu. Tādēļ prognozēšanas metode ir jāizvēlas, ņemot vērā prognozes intervālu, prognozes kļūdu un skaitļošanas izmaksas.

Šajā nodaļā tiek apskatīti sīkāk ARIMA/FARIMA un neironu tīklu modeļi no teorētiskā viedokļa. Izanalizējot ARIMA/FARIMA prognozēšanas modeļus, jāsecina:

- ARIMA procesi ir vienkārši un ērti piemērojami, taču šiem procesiem vienīgā problēma ir tā, ka ir sarežģīti pielāgot svāra koeficientus modelim ar empīriskām laika sērijām, t.i., optimizēts algoritms ir jāizveido pirms aktuālās optimizācijas un tam ir nepieciešami aktuāli datplūsmas dati.

- FARIMA modeļiem ir jāizvēlas liels parametru skaits, skaitļošanas laika ziņā tie aizņem ilgu laiku (vairākas dienas, atkarībā no tehniskajām iespējām), kamēr no datplūsmas trases ar garumu N tiek ģenerēta datplūsma ar garumu N . FARIMA modeļi var precīzi raksturot tīkla datplūsmu, kurai raksturīga tikai īstermiņa atkarība. Gadījumos, ja datplūsmai ir raksturīga ilgtermiņa atkarība, šie modeļi nav pielietojami datplūsmas prognozēšanai.

Analizējot neironu tīklu prognozēšanas modeļus, jāsecina:

- Lineārie tīkli sastāv no diviem slāņiem- ieejas un izejas. Slāņi ir izvietoti viens aiz otra. Neironu skaits ieejas slānī ir atkarīgs no tā, cik liels nolašu skaits ir analizējams un kādā veidā notiek apmācība un simulācija. Neironu skaits izejā veidojas atkarībā no tā, cik soļus uz priekšu prognozē. Lineārajos tīklos kā aktivēšanas funkciju izmanto lineāro funkciju.

- Vienvirziena tīkli sastāv no trim slāņiem- ieejas, slēptā un izejas slāņiem. Kā aktivēšanas funkcija tiek izmantota nelineāra – sigmoidālu funkcija. Izejas slānī aktivēšanas funkciju var būt gan lineāra, gan nelineāra.

- RBF tīkli sastāv no trim slāņiem- ieejas, slēptā un izejas slāņiem. Kā aktivēšanas funkcija tiek izmantota nelineāra – Gausa funkcija. Izejas slānī aktivēšanas funkcija ir lineāra. RBF tīklos apmācība notiek „bez skolotāja”, līdz ar to nav jādefinē mērķa vērtības.

Novērtējot prognozēšanas rezultāta kļūdu, jāsecina:

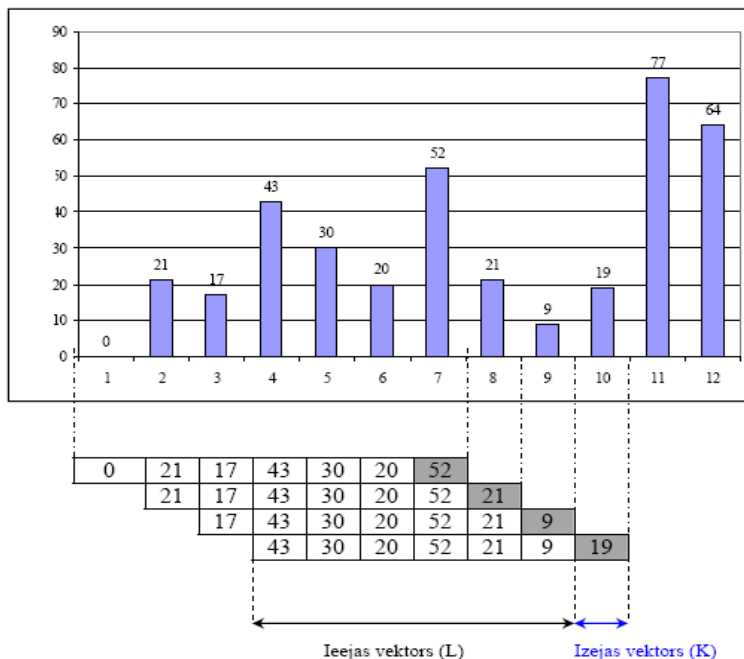
- Palielinot simulācijas soļu skaitu prognozēšanas procesā, VKK (vai jebkurš cits kļūdas novērtējums) būtiski nemainās. Veicot datplūsmas prognozēšanu, ir pietiekoši veikt 30 soļu simulāciju.

- Tā kā tiek veikta prognoze dažām datplūsmām ar dažādām nolašu vērtībām (maksimālās un minimālās), uzskatāmāk ir pielietot NVKK vai rezultātu procentuālo kļūdu, tā kā to lielumu neietekmē pašas nolašu vērtības.

5. Izmantotais prognozēšanas algoritms, prognozēšanas rezultāti

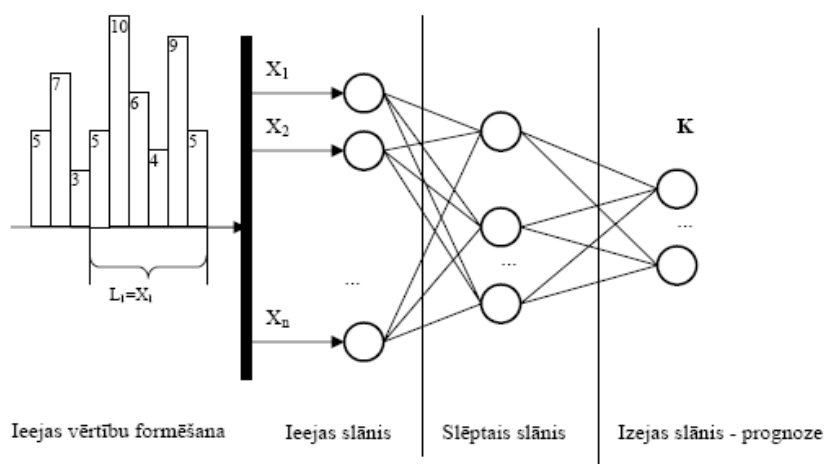
Promocijas darba gaitā tiek izmantots apmācības un nākotnes datu prognozēšanas algoritms, kas tiek piemērots iepriekšējā nodaļā aprakstītajiem neirona tīkla modeļiem - lineārais, vienvirziena vai radiālās bāzes tīkls. Tas satur 5 galvenos soļus:

1. Tiek iestādīta sākotnējā izmēģinājuma datu kopas klasifikācija - tiek iestādīts tīkla ieejas, slēptais un izejas slānis un tā vērtības. (skatīt 5.1. -5.2.attēlu)
 - ieejas slānis (pa kādiem vērtību vektoriem L datu masīvs tiks apstrādāts)
 - slēptais slānis (cik neironu tajā atradīsies)- attiecas tikai uz vienvirziena tīkliem
 - izejas slānis (prognozēšanas vērtības K)



5.1.att. Ieejas un izejas vektora formēšanas principi

5.1. attēlā ir parādīts, ka apmācība tiks realizēta ar tā saukto apmācības vektoru, kurš sastāv no divām daļām- ieejas vektora (ieejas vērtības) un izejas vektora (izejas vērtība). Šī apmācības vektora lielums ir atkarīgs no tīkla ieejas vērtībām (nolašu skaita) un iestādītajām izejas vērtībām (prognozēto soļu skaita). Eksperimenti tiks veikti ar prognozēto soļu skaitu $K=1, 10, 12, 24, 48, 72, 168, 336$ un 720 . Tā kā Freestats trasei (365d) autokorelācijai bija novērojams periodiskums $T=24$, tad tas tiks iekļauts prognozēšanas soļos. Pieņemot, ka $T=24h$, tad attiecīgi $K=24h$ (1 diena), $K=48$ (2 dienas), $K=72$ (3dienas), $K=168$ (1 nedēļa), $K=336$ (14 dienas) un $K=720$ (30 dienas jeb 1 mēnesis).



5.2.att. Neironu tīklu prognozēšanas shēma vienvirziena tīklos

2. tiek aprakstīts neirona tīkls (skatīt sīkāk 5.2. nodaļu).
3. pēc izveidotā tīkla notiek apmācība.
4. tiek definēti pamatprincipi nākotnes vērtību prognozēšanai. Nākotnes vērtību prognozēšanai tiek izmantotas Matlab Neural Network Toolbox piedāvātās opcijas neironu tīkla simulācijai.
5. tiek veikta prognoze un novērtēta tās precizitāte. Veiktās prognozes kļūda tiek novērtēta pēc procentuālās kļūdas.

5.2. Neironu tīklu modeļi

Šajā nodaļā šie modeļi tiek aprakstīti neironu tīklu modeļu ieejas, apmācības un simulācijas parametri, lai šie modeļi varētu veikt tīkla datplūsmas apmācību un prognozēšanu. Neironu tīklu modeļi tiek veidoti ar Matlab Neural Network Toolbox palīdzību.

5.2.1. Lineārie tīkli

Analizējot iegūtos prognozēšanas rezultātus ar lineārajiem tīkliem, tiek konstatēts:

- Reālajai datplūsmai ar īstermiņa atkarību (Freestats trase (365d)) labākie prognozēšanas rezultāti tiek uzrādīti pie $K=1$, $K=10$, $K=12$ un $K=720$. Mainot ieejas nogriežņa L garumu, prognozēšanas kļūda $E_{KL\text{UDA}}$ būtiski nemainās. Savukārt reālajam datplūsmām (Freestats trase (82d un 49d)) labākie rezultāti tika iegūti pie $K=1$ un $K \geq 24$. Pie lielākās laika soļu (K) izvēles tiek novērota tendence, ka prognozēšanas kļūda nepārsniedz 10% robežu. Tāpat pie $L=168$ un $L=720$ prognozēšanas kļūda nepārsniedz 10% (pie $K \geq 24$).
- Reālajai datplūsmai ar ilgtermiņa atkarību (Fotoblog trase) labākie prognozēšanas rezultāti tiek sasniegti pie $K=1$ un $K=24$. Deļ ilgā skaitļošanas laika eksperimentāli ir iegūti rezultāti tikai pie diviem ieejas vektoriem. Pie $K=12$ prognozēšanas kļūda sasniedz pat 160%, pie $K=48$ - 74%.
- Imitētajai datplūsmai ar īstermiņa atkarību (MMPP Freestats trase (365d, 82d un 49d)) labākie prognozēšanas rezultāti tiek uzrādīti pie $K=1$, $K=24$, $K=48$ un $K=72$. Mainot ieejas nogriežņa L garumu, prognozēšanas kļūda $E_{KL\text{UDA}}$ arī mainās- jo lielāks ieejas vektors L , jo mazāka prognozēšanas kļūda. Labāki prognozēšanas rezultāti, skatoties no ieejas vektora garuma, tiek iegūti pie $L=24$, $L=168$ un $L=720$.

- Imitētajai datplūsmai ar ilgtermiņa atkarību (MMPP Fotoblog trase) labākie prognozēšanas rezultāti tiek sasniegti pie $K=10$ un $K=24$. Dēļ ilgā skaitļošanas laika eksperimentāli ir iegūti rezultāti tikai pie diviem ieejas vektoriem.

Pamatojoties uz veikto analīzi, varam secināt:

- Reālajai datplūsmai ar īstermiņa atkarību:

- ieejas vektora L garums ir jāizvēlas tā, lai tas atbilstu $L=168$ (7dienas) vai $L=720$ (30 dienas), ja nolases ir izvēlētas ar laika soli 1h un autokorelācijas funkcijai ir novērojams cikliskums.

- prognozēšanas kļūda tiks sasniegta mazāka, ja prognoze tiks veikta pie $K \geq n \cdot 24$ ($n=1, 2, \dots$).

- Reālajai datplūsmai ar ilgtermiņa atkarību:

- ieejas vektora L garuma izvēlei, veicot prognozēšanu, nav būtiskas nozīmes.

- tā kā darbā apskatāmajai datplūsmai autokorelācijas funkcijas cikliskuma periods netika konstatēts, tad vislabākais rezultāts tiks atrasts pie $K=1$.

- Imitētajai datplūsmai ar īstermiņa atkarību:

- ieejas vektora L garums ir jāizvēlas tā, lai tas atbilstu $L=720$ (30 dienas), ja nolases ir izvēlētas ar laika soli 1h un autokorelācijas funkcijai ir novērojams cikliskums.

- prognozēšanas kļūda tiks sasniegta mazāka, ja prognoze tiks veikta pie $K \geq n \cdot 24$ ($n=1, 2, \dots$).

- Imitētajai datplūsmai ar ilgtermiņa atkarību:

- ieejas vektora L garuma izvēlei, veicot prognozēšanu, nav būtiskas nozīmes.

- prognozēšanas kļūda tiks sasniegta mazāka, ja prognoze tiks veikta pie $K=10$ un $K=24$.

5.2.2. Vienvirziena (arī daudzslāņu perceptronu) tīkli

Analizējot iegūtos prognozēšanas rezultātus ar vienvirziena tīkliem, tiek konstatēts:

- Reālajai datplūsmai ar īstermiņa atkarību (Freestats trase (49d)) labākie prognozēšanas rezultāti tiek uzrādīti pie $K=24$ un $K=48$. Savukārt reālajām datplūsmām (Freestats trase (365d un 82d)) labākie rezultāti tika iegūti pie $K=1$. Mainot ieejas nogriežņa L garumu, prognozēšanas kļūda $E_{kļūda}$ būtiski nemainās. Pie $K=10$ un $K=12$ prognozēšanas rezultāts sasniedz līdz pat 96% kļūdu, līdz ar to šo prognozēšanas apgabalu nevajadzētu izmantot.

- Imitētajai datplūsmai ar īstermiņa atkarību (MMPP Freestats trase (365d, 82d un 49d)) labākie prognozēšanas rezultāti tiek uzrādīti pie $K=1$. Palielinot ieejas nogriežņa L garumu, prognozēšanas kļūda $E_{kļūda}$ nedaudz samazinās. Pie $K=10$ un $K=12$ prognozēšanas rezultāts sasniedz līdz pat 159% kļūdu, līdz ar to šo prognozēšanas apgabalu nevajadzētu izmantot.

Pamatojoties uz veikto analīzi, varam secināt:

- Gan reālajai, gan imitētajai datplūsmai ar īstermiņa atkarību ieejas vektora L garumu vajag izvēlēties tā, lai tas atbilstu autokorelācijas funkcijas periodiskumam, bet ne mazāku par to. Vislabākais prognozēšanas rezultāts tiks sasniegts pie $K=1$.

- Palielinot neironu skaitu slēptajā slānī, tiek iegūta mazāka prognozēšanas kļūda. Ņemot vērā skaitļošanas laiku un rādītājus, kas iegūti mainot neironu skaitu slēptajā slānī, tiek rekomendēts izmantot 10 neironus slēptajā slānī.

5.2.3. Radiālās bāzes tīkli

RBF tīkli pēc savas būtības ir modeļi prognozēšanai „bez skolotāja”, līdz ar to skaitļošanas laiks ir ļoti ilgs. Ņemot vērā ļoti ilgo laiku, kas nepieciešamas vismaz vienas simulācijas izskaitļošanai, ir iegūti rezultāti tikai dažām datu trasēm- Freestats trasei (49d), Freestats trasei (82d), MMPP Freestats trasei (49d) un MMPP Freestats trasei (82d). Pārējām datu nolasēm bija nepieciešams liels datora atmiņas resursu nodrošinājums, kas nebija pieejams. Pēc eksperimentu rezultātiem var redzēt tendenci, ka palielinoties prognozēšanas soļu skaitam, palielinās prognozēšanas kļūda. Pēc iegūtajiem rezultātiem, var secināt, ka visoptimālākā prognoze tiek veikta 12 vai 24 soļiem uz priekšu. Šeit parādās likumsakarība ar autokorelācijas funkciju attiecīgi šīm datu trasēm (Freestats trasei (49d) un Freestats trasei (82d)), kur tika konstatēts periodiskums ar $T=24$. Tāpat pēc iegūtajiem rezultātiem var konstatēt, ka RBF tīklos nav nozīmes, cik lielu ieejas vektoru izvēlas prognozēšanas uzdevumam.

Balstoties uz veikto analīzi, galvenie secinājumi:

- RBF tīkliem nav būtiska ieejas vektora L izvēle.
- Reālajai datplūsmi, prognozējot ar RBF tīkliem, prognozēšanas kļūda ir lielāka nekā imitētajai slodzei.
- Gan reālajai, gan imitētajai datplūsmi ar īstermiņa atkarību prognozēšanas rezultāts tiks sasniegts labāks, ja $K \geq 12$.

5.3. Prognozēšanas rezultāta ticamības novērtējums

Neironu tīklu modeļos plaši tiek pielietoti dažādi aproksimācijas veidi. Daudzslāņu perceptronu tīkls, kurš sastāv no trim slāņiem apstrādā ieejas vektoru X no n neironiem, t.i., $\{x_1, x_2, \dots, x_n\}$ ieejas slānī, slēptajā slānī vektoru X no k neironiem un izejas slānī vektoru ar vienu neironu, ja tiek veikta prognoze vienam solim uz priekšu. Tādā gadījumā tīklam ir šāda vispārēja forma:

$$Y = f(X, \Theta) = f \left\{ b_0 + \sum_{j=0}^k v_j f_1 \left(\sum_{i=0}^n w_{ji} x_i + b_j \right) \right\} \quad (5.1)$$

kur Y raksturo izejas vērtības veidošanos;

Θ - apraksta vispārējo parametru telpu;

w_{ji} - apzīmē svaru koeficientus starp i -to ieejas neironu un j -to slēptā slāņa neironu;

v_j - apzīmē svara koeficientu starp izejas neironu un neironu j slēptajā slānī;

b_0 un b_1 -papildus nobīdes;

f_1 un f_2 - aktivēšanas funkcijas slēptajā un izejas slānī.

Ar Curve Fitting Toolbox palīdzību tiek izskaitļota ticamības josla pielāgošanas koeficientiem. Šajā gadījumā prognozes joslas un ticamības joslas ir viens un tas pats jēdziens, jo prognozētajām vērtībām tiek izskaitļotas ticamības joslas.

Ticamības josla pielāgošanās koeficientiem tiek uzdotas šādi:

$$C = k_{\text{fit}} \pm t\sqrt{S}, \quad (5.2)$$

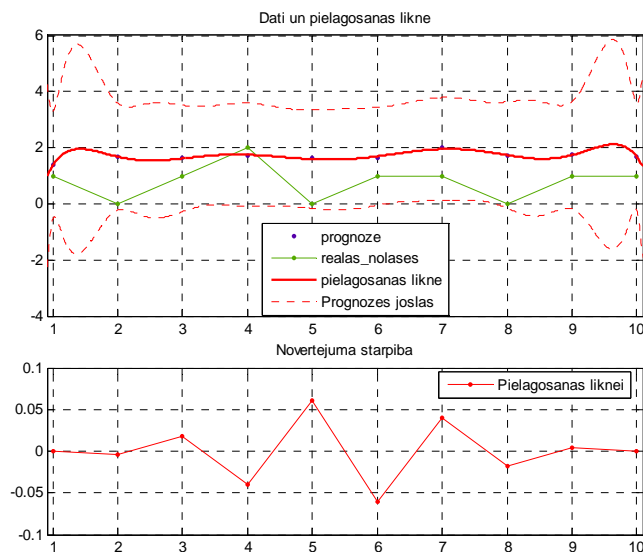
kur k_{fit} - pielāgošanās koeficienti;

t - inversā Stjudenta sadalījuma funkcija;

S tiek aprēķināts šādi:

$$S = \frac{1}{n} \sum_{i=1}^n \text{res}_i^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (5.3)$$

kur y_i - reālās nolases vērtība un \hat{y}_i - prognozētā nolases vērtība.



5.1.att. Prognozes joslas lineārajos tīklos pie ticamības pakāpes 95% Fotoblog trasei pie $K=10$ un $L=24$

Ticamības josla ar lineārajiem modeļiem Fotoblog trasei ir redzama 5.1.attēlā. 5.1.attēlā pielāgošanās likne tiek iegūta ar 8-tās kārtas polinomu. Pēc novērtējuma starpības var redzēt, ka atšķirība ir niecīga. Reālās laika nolases iekrīt ticamības joslā, līdz ar to var uzskatīt, ka rezultāti ir iegūti ar 95% ticamības pakāpi. Prognozes kļūdas novērtējums šim gadījumam bija $E_{kļūda} = 34.52$.

Veiktās prognozēšanas kļūdas ticamības novērtējums ir jāveic katram prognozēšanas gadījumam atsevišķi. Ja reālās nolases un prognozētās nolases (neatkarīgi no aprēķinātās prognozēšanas kļūdas) atrodas 95% ticamības intervāla joslās, rezultāts tiek uzskatīts par ticamu.

6. Neironu tīklu modeļa izvēles algoritms prognozēšanas rezultātam

6.1. Iegūto prognozēšanas rezultātu apkopojums

Veicot neskaitāmas simulācijas, ir iegūta šāda aina, kas atspoguļota 6.1.-6.3.tabulā (Tabulā katrā rindā labākie prognozēšanas rezultāti ir iezīmēti ar „Bold”, vislabākais rezultāts ir pasvītrots). Šeit netiek uzrādīti RBF tīklu prognozēšanas rezultāti, tā kā to skaits ir neliels.

6.1.tabula Vislabākā prognozēšanas kļūda lineārajos tīklos

Datplūsma	$K=1$	$K=12$	$K=24$	$K=72$	$K=168$	$K=720$
Freestats trase (49d)	11.903 ($L=1$)	1.60 ($L=168$)	0.025 ($L=168$)	0.015 ($L=720$)	0.515 ($L=720$)	3.924 ($L=168$)
Freestats trase (82d)	1.78 ($L=1$)	8.55 ($L=720$)	6.24 ($L=720$)	0.165 ($L=720$)	3.38 ($L=720$)	0.011 ($L=720$)
Freestats trase (365d)	0.22 ($L=1$)	4.282 ($L=336$)	25.75 ($L=24$)	30.273 ($L=168$)	23.76 ($L=168$)	5.073 ($L=720$)
MMPPFreestats trase (365d)	0.006 ($L=168$)	13.135 ($L=24$)	13.24 ($L=24$)	9.085 ($L=24$)	0.147 ($L=24$)	2.03 ($L=168$)
MMPP Freestats trase (49d)	0.617 ($L=720$)	4.613 ($L+720$)	1.186 ($L=168$)	1.198 ($L=720$)	-	-

MMPP Freestats trase (82d)	1.512 (L=720)	3.53 (L=720)	0.915 (L=168)	0.13 (L=1440)	-	-
Fotoblog trase	8.216 (L=168)	137.27 (L=168)	59.76 (L=24)	44.36 (L=168)	-	-
MMPP Fotoblog trase	102.47 (L=168)	27.93 (L=24)	0.624 (L=24)	-	-	-

6.1.tabulā var redzēt, ka prognozēšanas rezultāts katrai datplūsmas trasei ir savs un tiek sasniegts pie dažāda K un L. Jāpiezīmē, ka uzskatāmi var redzēt, ka MMPP Freestats trasei (49d), MMPP Freestats trasei (82d), Freestats trasei (82d) prognozēšanas kļūda ir salīdzinoši mazāka kā citos gadījumos.

6.2.tabulā arī var redzēt, ka prognozēšanas rezultāts katrai datplūsmas trasei ir savs un tiek sasniegts pie dažāda K un L. Jāpiezīmē, ka prognozēšanas rezultāti vienvirziena tīkliem ir lielāki nekā lineārajiem tīkliem.

6.2.tabula Vislabākā prognozēšanas kļūda vienvirziena tīklos

Datplūsma	K=1	K=10	K=12	K=24
Freestats trase (49d)	11.23 (L=98)	5.874 (L=49)	2.274 (L=98)	5.24 (L=168)
Freestats trase (82d)	1.206 (L=48)	24.18 (L=164)	15.692 (L=164)	12.197 (L=48)
Freestats trase (365d)	0.216 (L=24)	3.116 (L=168)	17.59 (L=24)	36.63 (L=24)
Mmpp Freestats trase	0.502 (L=24)	12.57 (L=168)	10.187 (L=168)	4.955 (L=168)
MMPPFreestats trase (49d)	2.517 (L=24)	70.67 (L=24)	79.293 (L=24)	0.175 (L=12)
MMPPFreestats trase (82d)	9.751 (L=12)	97.79 (L=168)	58.70 (L=164)	2.25 (L=48)

6.3.tabula Vislabākā prognozēšanas kļūda RBF tīklos

Datplūsma	K=1	K=10	K=12	K=24	K=48	K=72	K=168
Freestats trase (49d)	373.91	33.21	30.23	23.16	23.55	27.88	30.91
Freestats trase (82d)	170.91	12.48	9.34	19.58	20.56	20.68	18.32
MMPP Freestats trasei (49d)	22.2	4.75	2.82	1.94	6.02	8.18	10.77
MMPP Freestats trasei (82d)	39.13	22.71	6.56	10.33	6.62	10.56	11.24

Apkopojot iegūtos rezultātus, var izdalīt šādus būtiskus punktus par prognozēšanu:

1. Analizējamais datu apjoms. Tā kā pētījumos tiek izmantotas dažādas datu trases no tīkla aplikācijas līmeņa, tad visi secinājumi attiecas tikai uz interneta datplūsmu, kas ievākta aplikācijas līmenī (pieprasījums globālajam tīmeklim). Neironu tīklu modeļi savā būtībā ir ļoti sarežģīti, jo paredz sīku skaitļošanu. Jo lielāks būs analizējamo datu nolašu skaits, jo ilgāka

būs neironu tīkla rezultāta skaitļošana. Prognozēšanas uzdevumā ir svarīgi definēt, cik lielai ir jābūt ieejas informācijai, lai korekti varētu veikt prognozi. Tas tika realizēts, izvēloties dažādas datu trases (skatīt 2.1.tabulu).

2. Analizējamo datu raksturs. Kā viens no galvenajiem interneta datplūsmas raksturojošiem parametriem ir izvēlēts Hersta parametrs ($0.5 < H < 1$). Jo augstāks H un tuvāks vērtībai 1, jo pašlīdzīgāka (sarežģītāka) ir pētāmā datplūsma un novērojama ilgtermiņa atkarība. Prognozēšanas uzdevumā vislabākais variants būtu, ja H vērtība būtu tuvu 0.5 – ir novērojama īstermiņa atkarība.

3. Prognozēšanas rezultāts.

Prognozēšanas rezultāts tiek novērtēts pēc procentuālās kļūdas. Vislabākais prognozēšanas modelis ir lineārais tīkls, jo ar to tiek sasniegta vismazākā prognozēšanas kļūda.

6.3. Neironu tīklu modeļa izveidošana

Ieejas datu ievākšana, to novērtējums

Ieejas dati tiek analizēti, nosakot Hersta koeficientu. No tā tiek secināts, vai pētāmā datplūsma ir pašlīdzīga.

Kritēriju atlase, modeļa izvēle

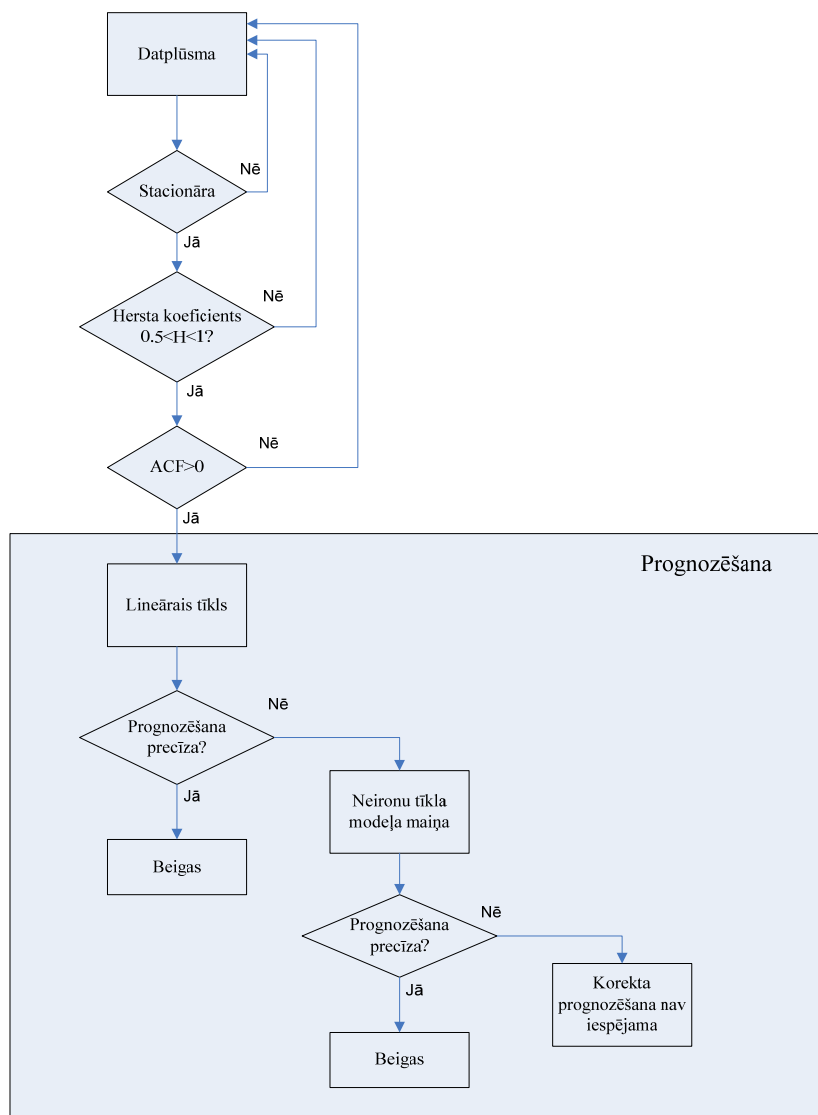
L izvēles kritēriji. Ieejas nogriežņa izvēle nosaka, pa kādiem cikliem izvēlētajam neironu tīklam apmācīties. Pētījumiem tiek izvēlēts L pēc diviem principiem- 1) pēc nejaušības un 2) pēc apkopojuma pakāpes nosakot datplūsmas pašlīdzības pakāpi pēc dispersijas- laika metodes.

Apkopotie rezultāti uzrāda, ka Lineārie tīkli ir vislabākie prognozētāji. Kopumā jārekomendē, ka L izvēle ir jāizvēlas pēc autokorelācijas funkcijas perioda - T , $7T$, $30T$ utt.

K izvēles kritēriji. Būtisks moments ir prognozēšanas nogriežņa K izvēle. Protams, ka vislabāk būtu, ja K būtu lielāks, taču, palielinoties K , pastāv varbūtība, ka palielināsies prognozēšanas kļūda. Šai sakarībā var secināt, ka ne vienmēr palielinoties K , prognozēšanas kļūda palielinās.

Promocijas darbā K vērtības tiek izvēlētas saistībā ar datplūsmas raksturu. Pieņemot, ka $T=24h$, tad attiecīgi $K=24h$ (1 diena), $K=48$ (2 dienas), $K=72$ (3dienas), $K=168$ (1 nedēļa), $K=336$ (14 dienas) un $K=720$ (30 dienas jeb 1 mēnesis).

Prognozēšanas modeļa izveidošana un modeļa pārbaude, iegūto rezultātu novērtēšana



6.1.att. Prognozēšanas procedūra

Secinājumi

Promocijas darba izstrādes laikā tika izvirzīti un veiksmīgi atrisināti šādi uzdevumi:
Tika izpētītas pašlīdzības efekta novērtēšanas metodes. Šim nolūkam tika izmantotas divas pašlīdzīgas datplūsmas noteikšanas metodes- dispersijas laika metode un R/S statistiskā metode.

Eksperimentāli, tika pierādīts, ka:

- Pētījumos izmantotās datu nolases (interneta datplūsma tīkla aplikācijas līmenī) ir pašlīdzīgas.
- Pašlīdzība ir novērojama datu nolasēm ar laika intervālu $T=1\text{sek}$ un $T=1\text{h}$.
- Pašlīdzības pakāpe palielinās, palielinoties analizē izmantoto nolašu skaitam.
- Reāla laika datu nolasēm Hersta koeficients pēc dispersijas- laika metodes mainās no 0.57-0.87.
- Imitētajām datu nolasēm Hersta koeficients pēc dispersijas- laika metodes mainās no 0.53-0.71.

- Reāla laika datu nolasēm Hersta koeficients pēc R/S statistiskās metodes mainās no 0.63-0.67.
- Imitētajām datu nolasēm Hersta koeficients pēc R/S statistiskās metodes mainās no 0.55-0.70.
- Pašlīdzības efekta novērtēšanas rezultāts pēc R/S statistiskās metodes būtiski atšķiras no rezultāta, kas iegūts pēc dispersijas- laika metodes.

R/S statistiskā metode uzrāda neprecīzus rezultātus datu trasēm, kas satur secīgas vērtības „0”, kā arī datu nolašu vērtībām ir tendence minimāli mainīties. Tas izskaidrojams ar to, ka starprezultātu kā R un S rezultāts ir 0, taču logaritmējot šo rezultātu attiecību R/S nav iespējams rezultāts, tā kā logaritms no 0 ir neiespējams.

Dispersijas –laika metodi var pielietot jebkurai datu nolasei. Pašlīdzības koeficienta novērtēšanas aprēķinos ir jāizmanto visas iespējamās apkopojuma pakāpes m , lai gala rezultāts būtu korekts.

Ir apskatīti dažādi neironu tīklu modeļi, kas var apmācīties un veikt prognozēšanu, balstoties uz vēsturiskajiem datiem. Promocijas darba izstrādē ir izpētīti trīs neironu tīklu modeļi- lineārie tīkli, vienvirziena tīkli un RBF tīkli. Visvienkāršākie ir lineārie tīkli, kas aizņem nelielu skaitļošanas laiku un uzrāda mazāku procentuālo kļūdu.

Ir veikta neironu tīklu prognozēšanas kļūdas ticamības joslu noteikšana. Prognozēšanas kļūdas ticamības novērtējums ir jāveic katram prognozēšanas gadījumam atsevišķi. Ja reālās nolases un prognozētās nolases (neatkarīgi no aprēķinātās prognozēšanas kļūdas) atrodas 95% ticamības intervāla joslās, rezultāts tiek uzskatīts par ticamu.

Tika pārbaudīta prognozēšanas kļūda reālajai datplūsmai un imitētai datplūsmai. Prognozēšanas kļūda reālajai un imitētajai datplūsmai būtiski atšķiras. Reālajai datplūsmai vislabākie prognozēšanas rezultāti tiek sasniegti pie $K \geq n \cdot 24$ ($n=1, 2, \dots$). Imitētajai datplūsmai vislabākie prognozēšanas rezultāti pārsvarā tiek sasniegti, prognozējot vienu soli uz priekšu ($K=1$).

Salīdzinātas dažādu neironu tīklu modeļu prognozēšanas iespējas. Vislabākie datplūsmas prognozēšanas modeļi ir lineārie tīkli. Veicot prognozi ar vienvirziena tīklu modeļiem, tiek iegūta lielāka prognozēšanas kļūda nekā ar lineārajiem tīklu modeļiem, bet RBF tīklu modeļi skaitļošana prasa lielus datora atmiņu resursus un lielu skaitļošanas laiku.

Izstrādātas neironu tīklu modeļa izvēles vadlīnijas prognozēšanas rezultātam. Ņemot vērā iepriekš eksperimentos iegūtos rezultātus, par prognozēšanu tika identificēti trīs būtiski faktori:

1. Analizējamais datu apjoms. Tā kā pētījumos tiek izmantotas dažādas datu trases no tīkla aplikācijas līmeņa, tad visi secinājumi attiecas tikai uz interneta datplūsmu, kas apkopota aplikācijas līmenī (pieprasījumu skaits globālajam tīmeklim). Neironu tīklu modeļi savā būtībā ir ļoti sarežģīti, jo paredz darbietilpīgu skaitļošanu. Jo lielāks būs analizējamo datu nolašu skaits, jo ilgāka būs neironu tīkla rezultāta skaitļošana. Prognozēšanas uzdevumā ir svarīgi definēt, cik lielai ir jābūt ieejas informācijai, lai korekti varētu veikt prognozi. Rekomendējoši būtu analizēt datu nolases, kuru skaits nepārsniedz 10 000, taču gadījumos, kad laika solis ir $T \leq 1$ sek, datu nolašu skaits būs lielāks un prognozes rezultāta skaitļošana ar neironu tīkliem aizņem ilgu laiku.

2. Analizējamo datu raksturs. Kā galvenie interneta datplūsmas raksturojošie parametri ir izvēlēti Hersta parametrs ($0.5 < H < 1$) un autokorelācijas funkcija. Prognozēšanas uzdevumā vislabākais variants būtu, ja H vērtība būtu tuvāk 0.5. Pēc autokorelācijas līknes var spriest, vai analizējamā datplūsma ir īstermiņa vai ilgtermiņa process. Tas ir būtiski, jo, izvēloties apmācības un prognozēšanas modeli īstermiņa un ilgtermiņa procesam, apmācības vektors un prognozēšanas vektora garums ir jāizvēlas atbilstoši autokorelācijas periodam, ja tāds pastāv.

3. Prognozēšanas rezultāts.

Prognozēšanas rezultāts tiek novērtēts pēc procentuālās kļūdas. Vislabākais prognozēšanas modelis ir lineārais tīkls, jo ar to tiek sasniegta vismazākā prognozēšanas kļūda. Plānojot tīkla datplūsmu, nav būtiski prognozēt kādu konkrētu vērtību konkrētā laika momentā, bet prognozēt noteiktā laika posmā iespējamās vērtības.

Ir izstrādātas šādas vadlīnijas, pēc kurām jāvadās prognozējot datu nolases ar neironu tīklu modeļiem:

- Lineārajos tīklos reālajai datplūsmai ar īstermiņa atkarību:
 - ieejas vektora L garums ir jāizvēlas tā, lai tas atbilstu $L=168$ (7dienas) vai $L=720$ (30 dienas), ja nolases ir izvēlētas ar laika soli $1h$ un autokorelācijas funkcijai ir novērojams cikliskums.
 - prognozēšanas kļūda tiks sasniegta mazāka, ja prognoze tiks veikta pie $K \geq n \cdot 24$ ($n=1, 2, \dots$).
- Lineārajos tīklos reālajai datplūsmai ar ilgtermiņa atkarību:
 - ieejas vektora L garuma izvēlei, veicot prognozēšanu, nav būtiskas nozīmes.
 - tā kā darbā apskatāmajai datplūsmai autokorelācijas funkcijas cikliskuma periods netika konstatēts, tad vislabākais rezultāts tiks atrasts pie $K=1$.
- Lineārajos tīklos imitētajai datplūsmai ar īstermiņa atkarību:
 - ieejas vektora L garums ir jāizvēlas tā, lai tas atbilstu $L=720$ (30 dienas), ja nolases ir izvēlētas ar laika soli $1h$ un autokorelācijas funkcijai ir novērojams cikliskums.
 - prognozēšanas kļūda tiks sasniegta mazāka, ja prognoze tiks veikta pie $K \geq n \cdot 24$ ($n=1, 2, \dots$).
- Lineārajos tīklos imitētajai datplūsmai ar ilgtermiņa atkarību:
 - ieejas vektora L garuma izvēlei, veicot prognozēšanu, nav būtiskas nozīmes.
- Lineārajos tīklos prognozēšanas kļūda tiks sasniegta mazāka, ja prognoze tiks veikta pie $K=10$ un $K=24$.
- Vienvirziena tīklos gan reālajai, gan imitētajai datplūsmai ar īstermiņa atkarību ieejas vektora L garumu vajag izvēlēties tā, lai tas atbilstu ar autokorelācijas funkcijas periodiskumu, bet ne mazāku par to. Vislabākais prognozēšanas rezultāts tiks sasniegts pie $K=1$.
- Vienvirziena tīklos palielinot neironu skaitu slēptajā slānī, tiek iegūta mazāka prognozēšanas kļūda. Ņemot vērā skaitļošanas laiku un rādītājus, kas iegūti mainot neironu skaitu slēptajā slānī, tiek rekomendēts izmantot 10 neironus slēptajā slānī.
- RBF tīkliem nav būtiska ieejas vektora L izvēle.
- Reālajai datplūsmai, prognozējot ar RBF tīkliem, prognozēšanas kļūda ir lielāka nekā imitētajai slodzei.
- Gan reālajai, gan imitētajai datplūsmai ar īstermiņa atkarību prognozēšanas rezultāts ar RBF tīkliem tiks sasniegts labāks, ja $K \geq 12$.

Aizstāvēšanai izvirzāmās tēzes

- Interneta datplūsmas pašlīdzības pakāpes novērtēšanai tīkla aplikācijas līmenī ir izmantojama dispersijas – laika metode.
- Neironu tīkla modeļos prognozēšanas uzdevumā ieejas vektors L ir jāizvēlas, ņemot vērā autokorelācijas funkcijas periodu (T) reizinātu ar kārtas reizinātāju n .
- Optimālais prognozēšanas intervāls ar neironu tīklu modeļiem pēc prognozēšanas kļūdas novērtējuma ir 1, 24 vai 720 soļi uz priekšu.

Literatūra

1. Bestavros A., Crovella M. E. Self-similarity in world wide web traffic: Evidence and possible causes// IEEE/ACM Trans. Networking.- 1997.- Vol. 5. - pp. 835–846.
2. Bodruzzaman M. Hurst's Rescaled Range (R/S) Analysis And Fractal Dimension Of Electromyography (EMG) Signal// IEEE Proceedings Southeastcon '91.- vol.2.- April 1991.- pp. 1121 – 1123.
3. Broersen, P.M.T. The prediction error of autoregressive small sample models// IEEE Transactions on Signal Processing, IEEE Transactions on Acoustics, Speech, and Signal Processing.- Volume 38.- May 1990.- pp. 858 – 860.
4. Cappe O., Pesquet J.C. Long Range Dependence and heavy tail modelling for teletraffic data// IEEE Signal Processing Magazine.- Volume 19.- Issue 3.- 2002.- pp. 14 – 27.
5. Chak, J.Q.-J., Zhuang, W. Connection admission control for indoor multimedia wireless communications// 48th IEEE Vehicular Technology Conference.- Volume 3.- 1998.- pp. 2570 – 2574.
6. Fowler. B.T., A Short Tutorial on Fractals and Internet Traffic. - 1999. - 14 pp.
7. Fowler. H., W. Leland Local Area Network Traffic Characteristics with Implications for Broadband Network Congestion Management//IEEE Journal on Selected Areas in Communications.- 9(7).- 1991.- pp. 1139-1149.
8. Huang, J., Le Ngoc T., Hayes, J.F. Broadband satcom system for multimedia services// IEEE International Conference on Communications.- Volume 2.- 23-27 June 1996.- pp.906 – 910.
9. Jonckheere E. A., Silvester J. A. Utilizing Neural Networks To Reduce Packet Loss In Self Similar Teletraffic Patterns.- 2003.-pp.1-5.
10. Lagrange, X., Godlewski, P. Teletraffic analysis of a hierarchical cellular network// IEEE 45th Vehicular Technology Conference.- Volume 2.- 25-28 July 1995.- pp.882 – 886.
11. Leland, W., M. Taqqu, W. Willinger, D. Wilson On the Self-Similar Nature of Ethernet Traffic (Extended Version)// IEEE/ACM Transactions on Networking. - 2(1). -1994. - pp. 1-15.
12. Li M., Jia W., Zhao W. A Whole Correlation Structure of Asymptotically Self-Similar Traffic in Communication Networks// Proceedings of the First International Conference on Web Information Systems Engineering.- Volume 1.- 2000.- pp. 480 – 485.
13. Li M., Jia W., Zhao W. Correlation from of Timestamp Increment Sequences of Self- Similar Traffic on Ethernet// Conference of Proceedings IEEE WISE.- 2000.
14. Lindgreen G., Holst U. Recursive Estimation of Parameters in Markov- Modulated Poisson Process// IEEE/ACM Transactions on Communications.- Vol. 43.- No. 11.- November1995.- pp. 2812-2820.
15. Mansfield G., Roy T.K., Shiratori N. Self-similar and fractal nature of Internet traffic data//15th International Conference on Information Networking.- 2001.- pp.227 – 231.
16. Park K., Wilinger W. Self Similar Network Traffic and Performance Evolutions// Wiley-Interscience.- August 2000.- p.558.
17. Pattavina A., Bregni S., Galimberti L. On the Analysis Tools of Traffic Fractality.- IEEE Global Telecommunications Conference.- GLOBECOM '03.- Volume 6.- 1-5 December 2003.- pp. 3211 – 3215.
18. Paxson V., Floyd S. Wide- Area Traffic: The failure of Poisson Modelling// IEEE/ACM Transactions on Networking.- Volume 3.- 1995.- pp. 226 – 244.

19. Peha J. M. Protocols Can Make Traffic Appear Self-Similar//IEEE/AMC/SCS Communication networks and distributed systems modeling and simulation conference.- January 1997.-pp.47-52.
20. Robert S., Le Boudee J.Y. A Markov Modulated Process For Self-Similar Traffic// Technical Report.- September 23 1995.- pp.1-14.
21. Rose O. Estimation of the Hurst Parameter of Long Range Dependent Time Series.- University of Wurzburg.- Research report series.- Feb.1996.-pp.1-15.
22. Selvakumar N., Tho Le-Ngoc Traffic Modelling in a Multi- media Environment// IEEE Transactions.- 1995.- pp.838-841.
23. Shah-Heydari, S., Tho Le-Ngoc MMPP modelling of aggregated ATM traffic// IEEE Canadian Conference on Electrical and Computer Engineering.- Volume 1.- 24-28 May 1998.- pp.129 – 132.
24. Subramanian, S.N., Tho Le-Ngoc Traffic modelling in a multi-media environment// Canadian Conference on Electrical and Computer Engineering.- Volume 2.- 5-8 Sept. 1995.- pp.838 – 841.
25. Tang D., Baker M. Analysis of a local –area wireless network// Proceedings MOBICOM'00. – Boston (USA). - August 2000. - pp.1-10.
26. Taqqu M.S., Teverovsky V. On Estimating the Intensity of Long Range Dependence in finite and Infinite Variance Time Series// “A Practical Guide to heavy tails: Statistical Techniques and Applications”.- Boston.- 1998.
27. Ulanovs P., Petersons E. Modeling methods of self similar traffic for network performance evaluation// Proceedings of RTU.- Telecommunications and Electronics.- Vol.7.- 2002.- pp.40-49.
28. Xueshi Y., Petropulu A.P., Pesquet J.-C. Estimating long-range dependence in impulsive traffic flows// Acoustics, Speech, and Signal Processing.- vol.6.- 2001.- pp. 3413 – 3416.
29. Yousef, S.Y., Strange, C.M., Schormans, J.A. ATM modelling: parameterisation of the 4-phase MMPP model for the superposed traffic sources// Proceedings IEEE ATM Workshop.- 25-28 May 1997.-pp.621 – 630.