

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
KOMPIUTERIJOS KATEDRA

Baigiamasis magistro darbas

Informacinių technologijų rizikos valdymo sistema

Atliko: 2 m. kurso, 10 grupės studentas

Linas Virbalas

Darbo vadovas:

doc. Algimantas Juozapavičius

Vilnius
2008

Turinys

Sutartiniai ženklai	3
Anotacija.....	4
Summary.....	5
Įvadas.....	6
1. Mokslinės srities apžvalga.....	8
2. Universalumas	9
3. Modelio principinė schema	11
4. Techninė realizacija.....	13
4.1. Architektūra.....	14
4.2. Koreliacijų radimo metodas.....	16
4.3. Neuroninių tinklų apsimokymo funkcija.....	17
4.4. Neuroninio tinklo architektūra	18
5. Pritaikymas	20
6. Eksperimentų rezultatai	22
Išvados	30
Literatūros sąrašas	31

Sutartiniai ženklai

Matematinėse išraiškose naudojamos raidės kilusios iš anglišku žodžių. Lengvesnio skaitymo dėlei stengtasi, kad jos turėtų prasmę: *I* nputs, *T* argets, *G* roup, *NN* (Neural Network).

$\|I\|$ - aibės *I* elementų skaičius.

i^* - išgalvotas arba sumodeliuotas dydis.

N - natūraliųjų skaičių aibė. Naudojama nurodyti indekso numeriui.

Daug kur naudojami sutrumpinimai:

Ap. – toliau tekste naudojamo simbolio apibrėžimas.

IT – informacinės technologijos.

IS – informacinės sistemos.

VP – vertybinis popierius (šio darbo kontekste – akcija).

IF – investicinis fondas.

OS – operacinė sistema.

Išsamumo dėlei skliausteliuose pateikiamos nuorodos į susijusias Matlab funkcijas ar raktinius žodžius, naudojamas fiksuoto pločio raidžių šriftas, pvz.:

(normc)

Anotacija

Šiuo darbu pristatoma sukurta sistema, kuria galima modeliuoti ir valdyti rizikas, kylančias iš IT, susijusias su IS nepasiekiamumu ar lėtu veikimu. Sistema realizuota pasitelkus neuroninius tinklus ir yra apmokoma sukaupta statistine informacija iš informacinių sistemų. Jai nurodoma, kurios statistinės informacijos laiko eilutes norima modeliuoti – t.y. kurios iš jų yra rizikos išraiška (serverių apkrovimas, IS atsakymo laikas ir pan.). Sistema pati nustato koreliuojančias statistines laiko eilutes, sugrupuoja susijusias ir kiekvienai grupei sukuria po modelį – apibendrina iki tol nežinomą priklausomybę tarp laiko eilučių pasitelkusi neuroninį tinklą. Kiekvienam iš tų modelių pateikus įtakojančių parametrų reikšmes, sistema sumodeliuoja rizikos parametro reikšmę. Eksperimentai parodė, jog sistema gali būti sėkmingai naudojama mišriame IT ūkyje ir geba modeliuoti įvairius IT bei IS komponentų parametrus, kurie sąlygoja rizikas.

Summary

By this work we present an IT risk management system, which is capable to model and manage risks that arise from IT which are related with IS downtimes and slow response times. The system is implemented by using a proposed neural network architecture as a heart of the modeling engine. It is trained with accumulated datasets from existing information systems. The user shows for the system which statistical data time series one needs to model – i.e. the one which represents the risk (like server load, IS response time, etc.). The system automatically determines correlated statistical time series, groups them and creates a separate model for each group – this model generalizes until then unknown relationship between time series by invoking neural network. The model then accepts values of the input parameters and the system models the value of the risk parameter. Experiments have shown that the proposed system can be successfully used in a mixed IT environment and can be rewarding for one who tracks IT risks coming from various IT and IS components.

Ivadas

Vis daugiau šiuolaikinių organizacijų ne tik naudoja informacines technologijas, tačiau IT joms yra ir įmonės veiklos užtikrinimo pamatas. Tai ypač ryškiai pastebima finansinėse, logistikos, bilietų rezervavimo, automobilių pramonės organizacijose ir t.t. Tokiose kompanijose nenumatyta ir nesuvaldyta IT problema gali sąlygoti organizacijai jos veiklos sustabdymą, reputacijos sugadinimą, milžiniškus finansinius nuostolius, bankrotą ar net žmonių gyvybes. Įmonių, IT departamentų ir projektų vadovams kyla pagrįstas poreikis įvertinti, valdyti ir sumažinti rizikas, kylančias iš IT ir IS.

Šiuo darbu pristatoma sistema, kuria galima modeliuoti ir taip valdyti (nuspėti) rizikas, kylančias iš IT, susijusias su IS nepasiekiamumu ar lėtu veikimu. Sistema realizuota pasitelkus neuroninius tinklus ir yra apmokoma sukaupta statistine informacija iš informacinių sistemų. Jai nurodoma, kurias statistinės informacijos laiko eilutes norima modeliuoti – t.y. kurios iš jų yra rizikos išraiška (serverių apkrovimas, IS atsakymo laikas ir pan.). Sistema pati nustato koreliuojančias statistines laiko eilutes, atskiria nesusijusias grupes ir kiekvienai grupei sukuria po modelį – apibendrina iki tol nežinomą priklausomybę tarp laiko eilučių pasitelkusi neuroninį tinklą. Kiekvienam iš tų modelių pateikus įtakojančių parametrų reikšmes, jis sumodeliuoja rizikos parametro reikšmę.

Sistema kurta tam, kad padėtų atsakyti į tokius ir panašius klausimus:

- kurie iš statistinės informacijos dydžių yra susiję? Kokia priklausomybė sieja juos?
- koks bus IS atsakymo laikas (angl. response time), jeigu pasikeis vartotojų, užklausių, tam tikrų vidinių scenarijų (pvz. pasirašomų el. sutarčių ar nupirktų prekių el. būdu) ir t.t. skaičius per laiko intervalą?
- kiek įvykdoma verslo operacijų, jeigu sistema veikia lėtai – kokie iš to seka nuostoliai?
- koks bus kieto disko ar diskų masyvo, procesoriaus ir kitų resursų apkrovimas prie tam tikro vidinių scenarijų skaičiaus per laiko intervalą?
- koks bus IS atsakymo laikas prie tam tikro serverio/kompiuterio resursų apkrovimo?
- kiek ilgai veiks nepertraukiamo maitinimo šaltiniai, dingus elektrai, prie tam tikro sistemos apkrovimo?
- ir t.t.

Be to siūloma sistema gali būti sėkmingai panaudota platesnėse IT rizikos valdymo metodikose ([SGF02], [Mos95], [JZ05], [PH04]), kaip įrankis, padedantis numatyti ir planuoti rizikas tiesiogiai kylančias iš IS nepasiekiamumo ar lėto veikimo.



1 iliustracija. Nenutrūkstama rizikos vadyba. Šaltinis: [PH04].

1. Mokslinės srities apžvalga

Šis darbas patenka į tokias mokslinių tyrinėjimų šakas:

1. rizikų, kylančių tiesiogiai iš IT bei IS naudojimo, valdymas ([SGF02], [JZ05], [DP04], [RL00], [CR99]),
2. laiko eilučių tarpusavio koreliavimo analizė ([Bou96]),
3. rizikų modeliavimas, pasitelkiant neuroninius tinklus ([Neu02], [EPR02], [YJZ06], [NYD07]).

Rizikų, kylančių iš IT bei IS problemų, nagrinėjimas dažniausiai sutinkamas abstrakčiuose darbuose (punktas nr. 1). Juose teikiamos įvairios rekomendacijos, kaip tas rizikas valdyti, įvedant formalizuotus procesus (darbo tvarkas) organizacijoje, veiklos tęstinumo planus, IS saugumo politikas [SA04] ir pan. Tuo tarpu darbų, kuriuose galima sutikti konkrečius matematiniais įrankiais pagrįstus siūlymus, yra retenybė (punktas nr. 3), o tarp tų, kurie yra, didesnis dėmesys jaučiamas programinės įrangos vystymo projektų rizikoms modeliuoti.

Šiuo darbu siekiama atkreipti dėmesį į tai, kad pasitelkiant konkrečius matematinius įrankius (kovariacijos matricą, neuroninį tinklą) ir duomenis, kurie yra pasiekiami organizacijos IT ūkyje, galima gana tiksliai modeliuoti techninės bei programinės įrangos veiklos sutrikimus ir taip užbėgti šiems nepageidajamiems įvykiams už akių. Ši problematika yra gana novatoriška ir kol kas sulaukus dar menko dėmesio mokslinių nagrinėjimų tarpe, tačiau vis didėjanti organizacijų priklausomybė nuo IT bei IS veiklos tęstinumo, neabejotinai didins susidomėjimą šia tema.

2. Universalumas

Kadangi sistema paremta neuroniniais tinklais ir yra apmokoma, jos rezultato kokybei labai didelę įtaką turi apmokymo duomenų aibė. Šią aibę sudaro statistinės informacijos laiko eilutės, gautos iš IS duomenų bazių agregatinių užklausų, vidinių protokolų (angl. log) analizės ir pan. (žr. 2 iliustraciją).

<code>SELECT COUNT(*), sutarties_data FROM klientai GROUP BY sutarties_data ORDER BY sutarties_data</code>	HTTP Access.log date, requests, pages
31,20071011	20071011, 1293472, 44882
28,20071012	20071012, 1154913, 43927
2, 20071013	20071013, 205663, 7160
38,20071015	20071015, 1292659, 50491
23,20071016	20071016, 1149523, 44954
49,20071017	20071017, 1190065, 43962
34,20071018	20071018, 1155982, 43667
28,20071019	20071019, 1038909, 39709
...	...

2 iliustracija. Statinės informacijos iš IS laiko eilučių pavyzdys. Kairėje – agregatinė užklausa iš duomenų bazės, dešinėje – HTTP serverio protokolo analizė.

Geriausius rizikos modeliavimo rezultatus sistema generuoja tada, kai yra apmokoma duomenimis, paimtais iš IS ir atitinkančiais tokius reikalavimus:

1. Maksimali duomenų įvairovė. Pvz.: jeigu mes norime numatyti, kada IS gali lėtai veikti, būtina pateikti bent kelis atvejus, kada taip buvo. Taip pat rezultatai bus tuo geresni, kuo bus didesnis įvairių kombinacijų tarp skirtingų laiko eilučių reikšmių skaičius. Praktikoje tam reikia surinkti pakankamai didelį istorijos laikotarpį.
2. Duomenys tarp dviejų kraštutinių verčių kinta pagal tam tikrą, tegu ir nežinomą, priklausomybę. Pvz.: jeigu didelis procesoriaus apkrovimas, nepertraukiamo maitinimo šaltinis senka greitai, jeigu mažas – lėtai; jeigu aktyvių IS vartotojų mažai arba daug, tačiau sudėtingų užklausų skaičius mažas – IS atsakymo laikas mažas, jeigu aktyvių

virtotojų mažai, o užklausų skaičius didelis – IS atsakymo laikas vidutinis, o jeigu daug ir aktyvių virtotojų, ir sudėtingų užklausų – IS atsakymo laikas didelis.

3. Duomenys pateikiami iš susijusių IS. T.y. sistema pritaikyta analizuoti vieno IS ūkio rizikoms. Jeigu bus pateikiami visiškai nesusijusių IS duomenys, sistema paprasčiausiai neras koreliacijų ir neras pagal ką apmokyti neuroninius tinklus.
4. Nėra duomenų anomalijų. Pvz.: serveris labai apkrautas, tačiau IS-os, kuri veikia tame serveryje, atsakymo laikas yra labai mažas – reiškiats arba trūksta dar vieno dėmens (gal IS veikia dviejuose serveriuose?), arba tuo metu buvo išimtinis atvejis - anomalija (pvz. IS buvo skubiai rankiniu būdu perkelta į kitą serverį). Tokie duomenys mažina modeliuojamų rezultatų kokybę, dėl to juos patartina identifikuoti ir panaikinti.

Svarbu paminėti, jog statistiniai duomenys bei jų koreliacijos neatskleidžia priežastinių ryšių. Galime paimti pavyzdį iš socialogijos: statistiniai tyrimai rodo, jog vyrai, kurie yra apsiženiję gyvena ilgiau. Tačiau iš šio statistinio teiginio negalima daryti išvados, kad vedybos yra ilgesnio vyrų gyvenimo priežastis! Priešingai, gali būti, kad tiesiog vyrams, kurie yra geros fizinės bei mentalinės būklės, lengviau susirasti žmoną []. Dėl to siekiant gerų sistemos modeliavimo rezultatų apibrėžiant esminių parametrų aibę svarbu pasirinkti dydžius, kurie yra kuo toliau pasekmės grandinėje. Pvz.: jeigu sistemai pateikiame tris parametrus – procesoriaus apkrovimą, kieto disko skaitymo/rašymo operacijų skaičių bei nepertraukiamo maitinio šaltinio išsikrovimo greitį – jie visi tarpusavy koreliuos, tačiau kaip modeliuotiną – esminį – parametą virtotojas turėtų nurodyti nepertraukiamo maitinio šaltinio išsikrovimo greitį, nes tai yra kitų dviejų parametrų pasekmė.

Gali kilti klausimas, gal ši sistema tikų ne tik IT srities rizikoms modeliuoti? Teoriškai taip - ji galėtų sėkmingai veikti su panašiais į aukščiau aprašytus duomenimis ir iš kitų sričių, ypač tokių, kur yra aiškus apibrėžtumas, pvz. mechanikos ar elektronikos procesų, hidro stebėjimų ir pan., tačiau kitos sritys nei IT yra už šio darbo apimties ribų.

3. Modelio principinė schema

Įsivaizduokime, kad vartotojas yra IT srities specialistas, kuris surinko statistinę informaciją pagal datą ir laiką iš įvairiausių tarpusavy tiesiogiai arba netiesiogiai susietų informacinių sistemų. Žinios apie egzistuojančius informacijos sąryšius pas jį yra menkos. Pažymėkime šią informaciją.

Ap. 1.1.

I – įvesties parametrų aibė.

$i \in I$ - aibės I elementas. $i = i(t)$ - laiko eilutė (statistinė informacija).

Tarkime, kad iš visos aibės statistinės informacijos laiko eilučių jį domina kelios “esminės” eilutės ir jis nori sužinoti su kuriomis kitomis laiko eilutėmis šios eilutės susijusios ir nori tas esmines eilutes modeliuoti.

Ap. 1.2.

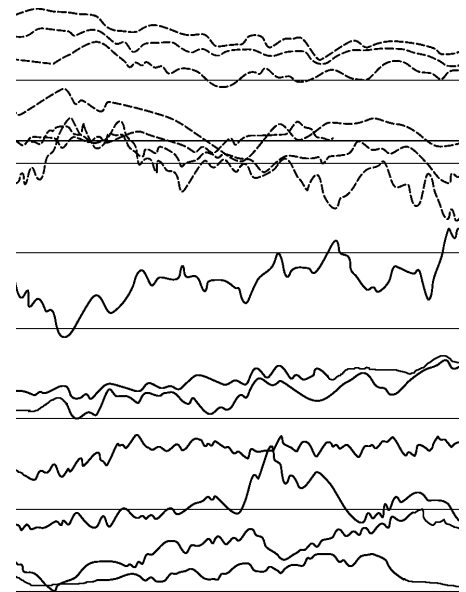
T - esminių parametrų aibė, kur $t \in T$ ir $T \subset I$.

Aprašomas modelis automatiškai atranda sąryšius tarp įvesties laiko eilučių (kartu ir esminių laiko eilučių) ir sugrupuoja jas.

Ap. 1.3.

$G_g = (I_g, T_g)$ - tarpusavy susijusių parametrų grupė su esminiais parametrais, kur $I_g \subset I$ ir $T_g \subset T$. Praktiškai dažniausiai elementų skaičius $\|T_g\| = 1$.

Identifikavęs grupes, modelis kiekvieną grupę apdoroja atskirai. Pagal grupėje esančias laiko eilutes yra apmokinamas neuroninis tinklas. Taip abstraguojama nežinoma priklausomybė tarp įvesties parametrų grupėje I_g ir esminių parametrų grupėje T_g .



3 iliustracija. Įvesties laiko eilučių pavyzdys. Punktyrais pažymėtos pasirinktos esminės laiko eilutės.

Ap. 1.4.

NN_g - apmokintas ([BM01], [Hea07]) neuroninis tinklas pagal mokymo duomenis $I_g \rightarrow T_g$.

Dabar vartotojas turi vertingą įrankį, su kuriuo gali modeliuoti jį dominančius esminius parametrus pagal sugalvotas įvesties parametų tam tikroje grupėje reikšmes.

Ap. 1.5.

$I_g^* = \{i_{g_1}^*, i_{g_2}^*, \dots, i_{g_n}^*\}$ - išgalvota įvesties parametų aibė neuroniniam tinklui NN_g . Čia $n = \|I_g\|$.

Svarbu paminėti, kad čia parametrai $i_{g_k}^*$ - jau skaliarai, o ne laiko eilutės.

$T_g^* = \{t_{g_1}^*, t_{g_2}^*, \dots, t_{g_m}^*\}$ - neuroninio tinklo NN_g sumodeliuotos (aproksimuotos) esminių parametų reikšmės įvedus I_g^* . Čia $m = \|T_g\|$, o $t_{g_k}^*$ - skaliarai.

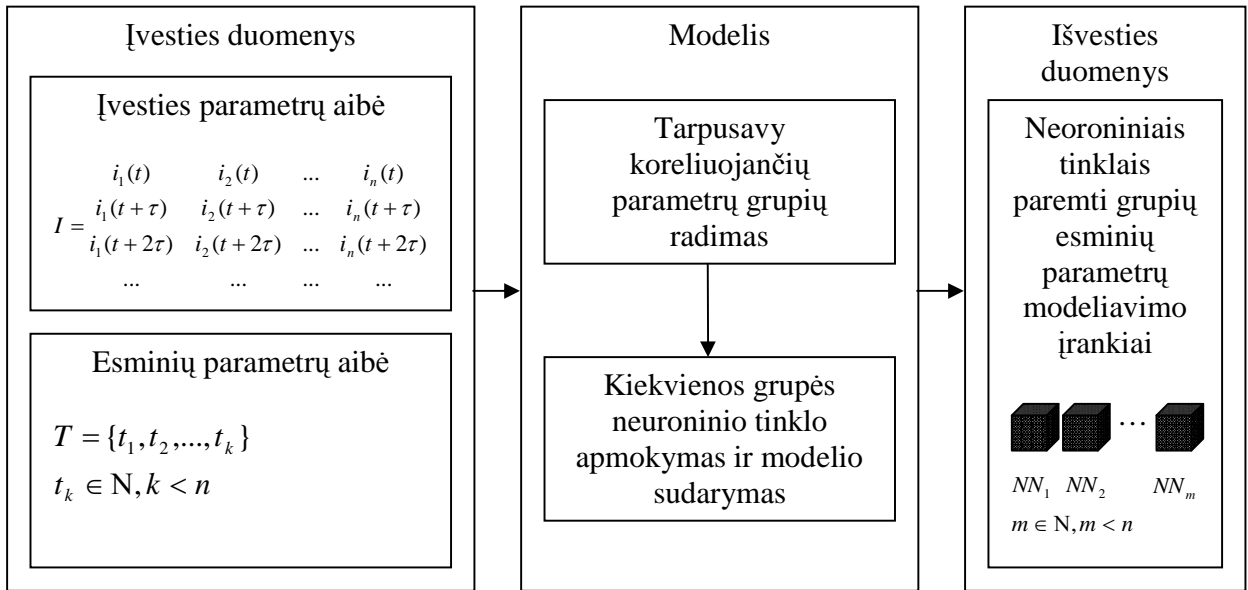
Į detalesnius klausimus, kokie metodai naudojami modelyje automatiniam įvesties parametų grupavimui, kokios struktūros neuroninis tinklas yra naudojamas ir kokia jam taikoma apmokymo funkcija, atsako skyrius “techninė realizacija”.

4. Techninė realizacija

Šiame skyriuje aprašoma sukurto įrankio techninė realizacija ir architektūra.

Sistemos įgyvendinimui pasirinktas Matlab programavimo paketas dėl plataus problematiką sprendžiančių įrankių pasirinkimo. Išsamumo dėlei toliau tekste bus pateikti ir naudojamų Matlab funkcijų pavadinimai.

4.1. Architektūra



4 iliustracija. Modelis ir jo įvesties bei išvesties duomenys.

Įvesties parametru aibę I (žr. ap. 1.1) vartotojas apibrėžia naudodamas matricą, kurios stulpelių skaičius atitinka aibės I parametru skaičių, o eilutės – laiką. T.y. matricos stulpelis yra laiko eilutė $i = i(t)$:

$$I = \begin{matrix} & i_1(t) & i_2(t) & \dots & i_n(t) \\ i_1(t+\tau) & i_1(t+\tau) & i_2(t+\tau) & \dots & i_n(t+\tau) \\ i_1(t+2\tau) & i_1(t+2\tau) & i_2(t+2\tau) & \dots & i_n(t+2\tau) \\ \dots & \dots & \dots & \dots & \dots \end{matrix}$$

Esminių parametru aibę T (žr. ap. 1.2) vartotojas apibrėžia tiesiog nurodydamas matricos I stulpelių numerius. Pvz.:

$$T = \{1,3,8\}$$

Tada modelis ieško tarpusavy koreliuojančių įvesties laiko eilučių ir sudaro grupes $G_g = (I_g, T_g)$ (žr. ap. 1.3). Technologiškai grupė taip pat aprašoma tiesiog matricos I stulpelių numeriais. Suprantama, mus domina tos grupės, kuriose yra elementų iš aibės T . Pvz.:

$$G_1 = \{6,7,3\}$$

$$G_2 = \{2,4,5,1\}$$

Pasaulyje yra įvairių metodų skirtų koreliuojančių laiko eilučių paieškai atlikti. Šiame modelyje naudojama koreliacijos koeficientų matrica, kuri gaunama iš kovariacijos matricos. Detalesnis aprašymas ir argumentacija pateikta poskyryje “koreliacinis modelis”.

Tada modelis apdoroja kiekvieną grupę G_g atskirai:

1. sunormuoja grupės matricos stulpelius (`normc`, `normr`) – to reikia, nes šiame modelyje naudojama neuroninio tinklo apmokymo funkcija, kuri priima būtent sunormuotus duomenis (žr. poskyrį “neuroninių tinklų apsimokymo funkcija”);
2. sunormuotą matricą pateikia paruoštam neuroniniam tinklui, iš kurios receptoriams pateikiamos laiko eilutės $i \in I$, o taikinyis yra laiko eilutės $t \in T$. Tada paleidžiama neuroninio tinklo apsimokymo funkcija ir gaunamas šią grupę G_g atitinkantis neuroninis tinklas NN_g - įrankis grupės esminių parametrų T_g modeliavimui.

4.2. Koreliacijų radimo metodas

Sistema kurta taip, kad jai galima būtų pateikti daugybę statistinių įvesties parametru apsimokymui – daugybę laiko eilučių iš IS ūkio. Rankiniu būdu identifikuoti, kurie iš tų parametru yra susiję būtų painu ir neracionalu, dėl to sistema tai atlieka automatiškai.

Eksperimentai parodė, kad įvesties parametrai, kurie yra tarpusavy susiję, turi gan stiprų tiesinį sąryšį. Pvz.: sutarčių pasirašytų IS-oje per dieną kiekis yra tam tikru stiprumu tiesiškai susijęs su serverio apkrovimo dienos vidurkiu. Dėl šios priežasties susijusių parametru paieškai pasitelkta plačiai naudojama koreliacijos koeficientų apskaičiavimo funkcija (`corrcoef`) [Bou96]. Jai pateikus įvesties parametru reikšmių matricą I , funkcija grąžina normalizuotą tiesinio sąryšio stiprumo įvertį tarp kiekvieno iš matricos įvesties parametru:

$$R(i, j) = \frac{C(i, j)}{\sqrt{C(i, i)C(j, j)}}, \text{ čia } C(i, j) - \text{kovariacijos matrica.}$$

Tada yra pasirenkami akivaizdžiai koreliuojantys parametrai ir taip sudaromos anksčiau minėtos grupės G_g . Tiesa, paliekamos tik tos grupės, kurios turi nors vieną esminį parametru $t \in T$.

4.3. Neuroninių tinklų apsimokymo funkcija

Neoriniai tinklai dažnai laikomi universaliais funkcijų aproksimatoriais, nes teoriškai bet kokia tolydi funkcija gali būti aproksimuojama iki reikiamo tikslumo lygio didinant neuronų skaičių tinkle [Die03]. Tačiau praktikoje dažniausiai labiau reikia ne tikslaus aproksimavimo, o apibendrinimo tiems atvejams, kurie nebuvo pateikti apmokymo aibėje. Tinklas, kuris yra per daug sudėtingas ir turi per daug neuronų, lyginant su apmokymo aibės sudėtingumu, yra linkęs blogai apibendrinti (angl. *overfitting*) – tokiais atvejais yra tiksliai aproksimuojamos žinomos reikšmės, tačiau pateikus tinklui nematytą atvejį, jis gražina neteisingą reikšmę [Sar01]. Kadangi šio modelio tikslas yra suteikti vartotojui įrankį, kuriuo jis galėtų numatyti ir valdyti rizikas, kylančias iš IT sistemų, yra labai svarbu, kad būtų maksimaliai teisingai aproksimuojamos iki šiol nematytos neuroniniam tinklui reikšmės. Švelnus perėjimas nuo vienos žinomos reikšmės iki kitos yra būtinas, nes, jeigu buvo pateikti tinkami apmokymo duomenys, tai būtent toks perėjimas geriausiai atitiks laukiamą rezultatą (žr. skyrių “universalumas”). Pasaulyje siūlomi bent du būdai, kurie sprendžia blogo apibendrinimo (angl. *overfitting*) problemą: ankstyvas apmokymo sustabdymas (angl. *early stopping*) ir Baeso apmokymas [KY96] (Bayesian regularization). Lyginant su ankstyvo sustojimo būdu, Baeso apmokymas sugeba geriau apibendrinti - ypač tai pastebima su mažesnėmis apmokymo duomenų aibėmis, taip pat tokiomis, kur yra šiek tiek triukšmo [DL20]. Kaina, kurią tenka už tai mokėti – greitis [DBH07], tačiau tai šioje sistemoje nėra esminis veiksnys, dėl to pasirinktas būtent Baeso apmokymas (`trainbr`).

4.4. Neuroninio tinklo architektūra

Nežinomai priklausomybei nustatyti tarp įvesties parametrų ir norimo modeliuoti esminio parametro pasitelktas prižiūrimo mokymo neuroninis tinklas be ciklų (supervised learning feedforward.network). Ieškota tokios neuroninio tinklo architektūros, kuri sėkmingai aproksimuotų nuo 2 iki 10 įvesties parametrų vienoje grupėje. Didesnis parametrų skaičius grupėje laikytas neracionaliu, nes tada pats modeliavimas vartotojui tampa per daug painus.

Kadangi vieningos ir patikimos metodologijos, kaip pasirinkti neuroninio tinklo architektūra bei neuronų skaičių nėra [Sar01], buvo pasitelktas eksperimentavimas. Vieno paslėpto sluoksnio neuroniniai tinklai rodė neblogus rezultatus, tačiau dėl reikalingo didelio kiekio neuronų viename sluoksnyje labai lėtai apsimokė ir dėl to nebuvo įmanoma juos tinkamai išbandyti (žr. eksperimentus nr. 10, 11, 12). Kitos architektūros gerai veikė su daugiau nei 5 įvesties parametrais, tačiau prastai su mažiau (žr. eksperimentą nr. 6). Pati optimaliausia architektūra tiek kokybės prasme (sėkmingai apsimokanti su įvairiais parametrų kiekiais), tiek greičio pasirodė esanti sudaryta iš trijų paslėptų sluoksnių po 15, 7 ir 15 neuronų juose atitinkamai (žr. eksperimentą nr. 8).

Nr.	Neuroninio tinklo architektūra (neuronų skaičius sluoksniuose, atitinkamų sluoksnių perkelties (angl. transfer) funkcijos ir apmokymo metodas	SSE (sum squared error) po X epochų su 7 (jei kitaip nenurodyta) įvesties parametrais
1.	[10,10,1],{'tansig','tansig','purelin'}, 'trainbr'	Epoch 300/300, SSE 0.000560847
2.	[12,10,1],{'tansig','tansig','purelin'}, 'trainbr'	Epoch 300/300, SSE 0.00147872
3.	[15,15,1],{'tansig','tansig','purelin'}, 'trainbr'	Epoch 300/300, SSE 0.000855001
4.	[7,7,1],{'tansig','tansig','purelin'}, 'trainbr'	Epoch 300/300, SSE 8.79522e-006/0 (tik vieną kartą taip gerai)
5.	[5,5,1],{'tansig','tansig','purelin'}, 'trainbr'	Epoch 300/300, SSE 0.00086876/0
6.	[7,7,7,1],{'tansig','tansig','tansig','purelin'}, 'trainbr'	Epoch 300/300, SSE 0.000286456 Epoch 500/500, SSE 3.67415e-005 Epoch 1834/5000, SSE 0.000830797 Epoch 2989/5000, SSE 4.08624e-009 (479.609000 s)

		Su 4 įvesties parametrais: Epoch 465/5000, SSE 0.00144585
7.	[4,4,4,1],{'tansig','tansig','tansig','purelin'},'trainbr'	Su 4 įvesties parametrais: Epoch 465/5000, SSE 0.00170492
8.	[15,7,15,1],{'tansig','tansig','tansig','purelin'},'trainbr' <u>(greitas)</u>	Su 6 įvesties parametrais: <u>Epoch 18/5000, SSE 0.000743112</u> Epoch 500/5000, SSE 0.00132527 <u>Epoch 100/5000, SSE 4.86856e-005</u> Su 4 įvesties parametrais: <u>Epoch 300/5000, SSE 6.70256e-005</u> <u>Epoch 500/5000, SSE 6.09294e-005</u> Su 2 įvesties parametrais: <u>Epoch 1834/5000, SSE 0.00147841</u>
9.	[50,50,1],{'tansig','tansig','purelin'},'trainbr'	Epoch 10/300, SSE 0.00568961
10.	[100,1],{'tansig','purelin'},'trainbr' (lėtas)	Epoch 48/50, SSE 0.000375572 Epoch 50/50, SSE 0.000454926 Epoch 265/500, SSE 0.000255758
11.	[200,1],{'tansig','purelin'},'trainbr' (lėtas)	Epoch 15/50, SSE 4.531e-005 Epoch 18/50, SSE 0.000118878
12.	[300,1],{'tansig','purelin'},'trainbr' (lėtas)	Epoch 13/50, SSE 2.60478e-006 Epoch 50/50, SSE 0.000139524

1 lentelė. Skirtingos neuroninio tinklo architektūros eksperimentų rezultatų ištrauka.

Nustačius optimalią architektūrą [15,7,15,1] eksperimentai buvo tęsiami. Išbandžius įvairias įvesties parametų aibes, pastebėta, kad net ir IT parametų modeliavimo sritis yra pakankamai plati ir ta pati neuroninio tinklo architektūra vienoms aibėms gali generuoti tikslesnius rezultatus negu kitoms (žinoma, tai priklauso ir nuo pačių duomenų kokybės). Todėl, kai neuroninis tinklas yra taikomas konkrečiai duomenų sričiai, jei yra galimybė, rekomenduotina atlikti pirminę analizę, išbandant įvairesnes neuroninių tinklų struktūras, tokias kaip: [10,5,10,1], [15,10,15,1], [15,20,15,1] ir pan. Pirminė analizė užima laiko, tačiau jos pasekoje gali pasirodyti, kad tenkinantiems rezultatams pasiekti, galima naudoti mažiau neuronų tinkle ir taip pagreitinti algoritmą, o tam tikrais atvejais net patikslinti modeliavimą (žr. lentelę nr. 3 skyriuje „eksperimentų rezultatai“).

5. Pritaikymas

Ši sistema kurta, kad galėtų būti pritaikyta organizacijose, kuriose yra svarbu valdyti rizikas kylančias iš IS nepasiekiamumo arba lėto veikimo. Lentelėje nr. 2 aprašyti keli sistemos panaudojimo scenarijai. Juos, suprantama, galima naudoti visus kartu.

Nr.	Aprašymas	Įvesties parametru aibė I	Esminių parametru aibė T
1	IS veikimo laikotarpio modeliavimas dingus elektrai, kai IS veikia dėka nepertraukiamo maitinimo šaltinio.	IS naudojimo intensyvumo parametrai (aktyvių vartotojų, operacijų skaičius ir t.t.) dingus elektrai.	Atitinkamas nepertraukiamo maitinimo šaltinio išsikrovimo greitis.
2	IS atsakymo laiko modeliavimas pagal techninės įrangos apkrovimą.	Techninės įrangos (procesoriaus, diskų masyvo, operatyviosios atminties ir t.t.) apkrovimo parametrai.	Atitinkamas IS atsakymo laikas.
3	IS atsakymo laiko modeliavimas pagal IS naudojimo intensyvumą.	IS naudojimo intensyvumo parametrai (žr. scenarijų 1).	Atitinkamas IS atsakymo laikas.
4	Nesėkmingai įvykdytų operacijų kiekio modeliavimas pagal IS atsakymo laiką, kai IS veikia nekorektiškai.	Nesėkmingai įvykdytų operacijų kiekis.	Atitinkamas IS atsakymo laikas.
5	Techninės įrangos apkrovimo modeliavimas pagal IS naudojimo intensyvumą.	IS naudojimo intensyvumo parametrai (žr. scenarijų 1).	Atitinkamas techninės įrangos (žr. scenarijų 2) apkrovimas.
6.	IS atsakymo laiko modeliavimas pagal ryšių apkrovimą, kai ryšiai tampa lėtinančiu veiksmiu.	Ryšio kanalų apkrovimo matavimai.	Atitinkamas IS atsakymo laikas.
7.	Ryšio kanalo apkrovimo modeliavimas pagal IS-ų naudojimo intensyvumus.	IS-ų naudojimo intensyvumo parametrai (žr. scenarijų 1).	Ryšio kanalo apkrovimo matavimai.

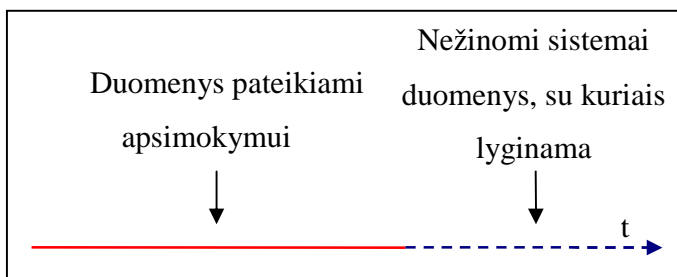
2 lentelė. Sistemos panaudojimo scenarijai.

Vienas didesnių šios sistemos privalumų yra tas, kad ji padeda suvaldyti rizikas dideliame IT ūkyje. Ten, kur IS įvairiais ryšiais viena su kita susijusios, kur architektūrinis aprašymas yra didelis, o gal tokio net nėra, šiai IT rizikų valdymo sistemai užtenka pateikti tik istorinius stebėjimo duomenis – tuo tarpu ryšių tarp įvairių duomenų radimu, grupavimu ir nežinomų priklausomybių tarp skirtingų parametrų apibendrinimu sistema pasirūpina pati.

6. Eksperimentų rezultatai

Šiame skyriuje pateikiami eksperimentų rezultatai, iliustruojantys atlikto tyrimo esmę. Kiekvienas eksperimentas pateikiamas atskirame puslapyje. Eksperimentams naudoti produkcinės aplinkos duomenys iš finansų srities organizacijos produkinių IS (eksperimentas nr. 1 - 5), skirtingų procesų, veikiančių vienoje OS-oje, įtaka bendram sistemos apkrovimui (lentelė nr. 3) ir k.t.. Eksperimentuota su skirtingomis įvesties parametru grupėmis.

Eksperimento metu sistemai pateikiama tik dalis turimų statistinių duomenų, likusi dalis jai lieka nežinoma (ilustracija nr. 5). Tada pagal likusią dalį yra modeliuojama. Tokiu būdu gaunamas objektyvus būdas su kuo palyginti sistemos sumodeliuojamas reikšmes. Prie kiekvieno eksperimento



5 iliustracija. Modeliavimo rezultatų tikrinimas.

pateikiami grafikai, kurie tai pavaizduoja. Grafike atkreiptinas dėmesys į lūžio momentą, kai peržengiama apmokintų situacijų skaičius – nuo to momento sistema prognozuoja iki tol jai nežinomas situacijas. Kartais sistema nesugebėdama sumodeliuoti situacijos grąžina neigiamą reikšmę – tai geriau negu klaidingai sumodeliuota reikšmė.

Panagrinėkime eksperimentus nr. 1 – 5. Eksperimentui pateikiama tokia įvesties parametru aibė:

$I =$

20071011	0	1	31	0	0	0	3304	49	1293472	44882
20071011	8	2	31	1	0	58	3304	49	1293472	44882
20071011	9	8	31	22	0343		3304	49	1293472	44882
20071011	10	15	31	37	0301		3304	49	1293472	44882
20071011	11	11	31	14	0339		3304	49	1293472	44882
20071011	12	10	31	7	0393		3304	49	1293472	44882
20071011	13	10	31	17	0	66	3304	49	1293472	44882
20071011	14	3	31	3	0	0	3304	49	1293472	44882
20071011	15	4	31	6	0	0	3304	49	1293472	44882
20071011	16	7	31	5	0	0	3304	49	1293472	44882
20071011	17	8	31	1	0	0	3304	49	1293472	44882
20071011	18	1	31	0	0	0	3304	49	1293472	44882
20071011	20	6	31	0	0	0	3304	49	1293472	44882
20071011	21	4	31	0	0	0	3304	49	1293472	44882

20071011	22	1	31	0	0	0	3304	49	1293472	44882
20071011	23	7	31	0	0	0	3304	49	1293472	44882
20071012	0	1	28	0	0	0	2909	62	1154913	43927
20071012	1	1	28	0	0	0	2909	62	1154913	43927
20071012	8	7	28	0	0	73	2909	62	1154913	43927
...										

Ją sudaro šie parametrai atitinkamai: data, valanda, per dieną svetainėje pateikti investicinių fondų (toliau – IF) pavedimai, pasirašytos sutartys, pateikti vertybinių poierių (toliau – VP) pavedimai ofise, dėl techninių kliūčių užstrigę VP pavedimai, VP sandoriai biržoje, vartotojų prisijungimų skaičius į svetainę, nesėkmingų prisijungimų skaičius į svetainę, užklausų skaičius į svetainę, atidarytų puslapių skaičius svetainėje.

Eksperimente nr. 1 esminis parametras buvo nurodytas “atidarytų puslapių skaičius svetainėje“ ir sistema automatiškai identifikavo grupę iš 7 narių, kuri įtakoja šį parametą. Kaip matyti iš rezultatų, sistema pakankamai tiksliai sugebėjo prognozuoti jai nežinomą laikotarpį.

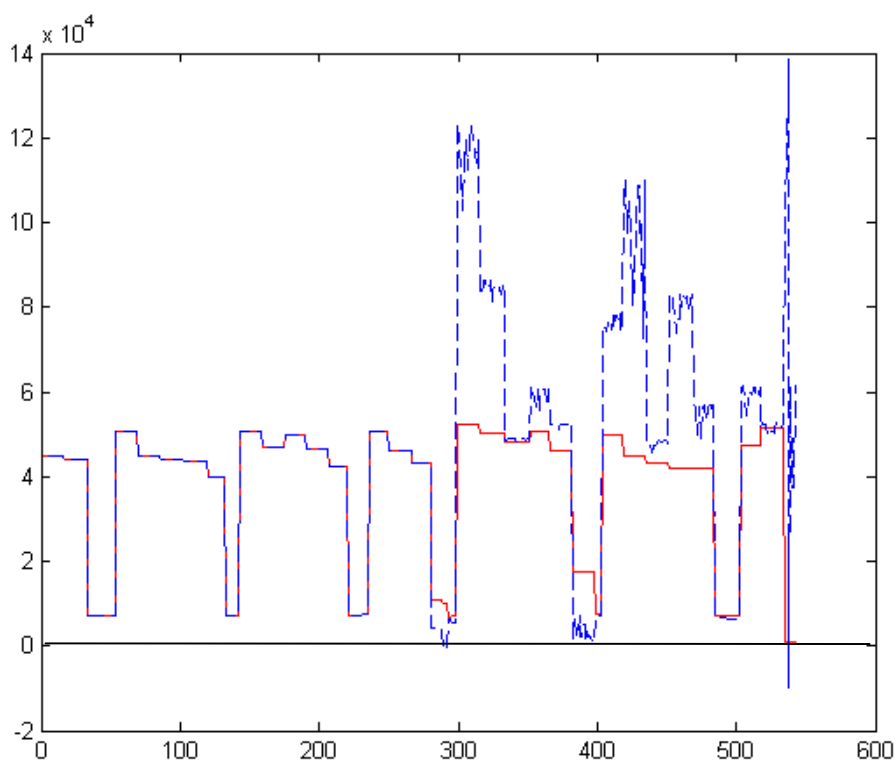
Kituose eksperimentuose įvesties parametų grupė buvo tyčia sumažinta iki 5 narių. Eksperimente nr. 2 buvo drastiškai padidinta nežinomų sistemai duomenų aibė beveik iki pusės visų duomenų dalies. Šįkart nors modeliuojamo esminio parametro reikšmės paklaida buvo didelė, tačiau tendencijas sistema sugebėjo prognozuoti patikimai. Sumažinus nežinomų duomenų kiekį iki ketvirtadalio (eksperimente nr. 3), matyti, kad sistema pradėjo žymiai tiksliau modeliuoti.

Stebint šių ir kitų atliktų eksperimentų rezultatus matyti, jog sistema sugeba gan tiksliai nustatyti susijusių parametų grupes - sumažinus ar pakeitus grupės narius, rezultatai dažniausiai pablogėdavo. Taip pat sistema sugeba modeliuoti ir pagal jai anksčiau nematytas įvesties parametų reikšmes. Tačiau, suprantama, jeigu grupės įvesties parametų reikšmių kombinacija sistemai yra visiškai nematyta, modeliavimas nebus sėkmingas. Dėl šios priežasties labai svarbu sukaupti kuo įvairesnių duomenų atvejų apmokymui (žr. skyrių „universalumas“).

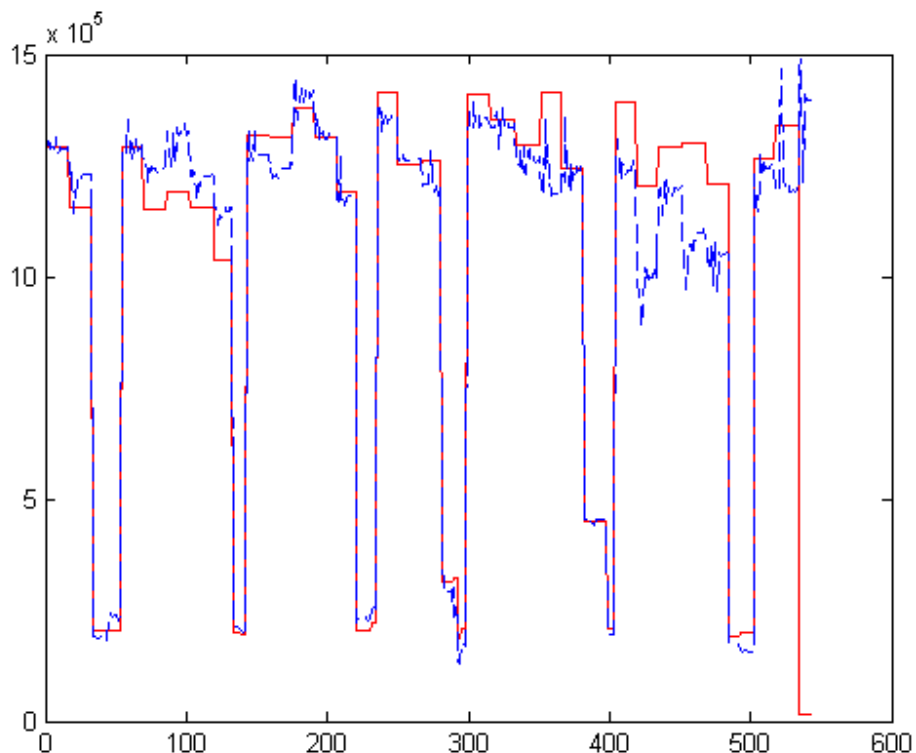
Taigi nors šiame darbe eksperimentai atlikti su duomenimis, kurių kiekis buvo, palyginus, mažas, galima daryti išvadą, kad naudojant sistemą tikroje organizacijos veikloje, tikėtina, kad galima pasiekti žymiai geresnius rezultatus, nes apmokymo aibė pasipildytų kasdieniais duomenimis iš įvairių IS. Taip sistema kiekvieną dieną mokintūsi ir darytūsi vis tikslesnė bei universalesnė.

Neuroninio tinklo apmokymo kokybė po 500 epochų		Viso situacijų: 543 (2007-10-11 - 2007-11-21)	
Experimentas nr. 1			
Įvesties parametrai (kiekis per dieną)		Esminis parametras (kiekis per dieną)	
1. Svetainėje pateikti IF pavedimai 2. Pasirašytos sutartys 3. VP pavedimai ofise 4. VP sandoriai biržoje 5. Prisijungimų skaičius į svetainę 6. Nesėkmingų prisijungimų skaičius į svetainę 7. Užklausų skaičius į svetainę		->	Atidarytų puslapių skaičius svetainėje
Apmokymo laikas (s):	136.812	Apmokintą situacijų:	403 (iki 20071111)
Žinomų situacijų modeliavimo kokybė		Nežinomų situacijų modeliavimo kokybė	
SSE (sum squared error):	2.10493e-008/0	Nežinoma dalis:	<u>25.78%</u>
Minimali klaida:	5.08047416636348e-009	Minimali klaida:	2.99796386604712e-008
Vidutinė klaida:	4.82319999573711e-006	Vidutinė klaida:	0.00861507359732967
Maksimali klaida:	3.27610067399065e-005	Maksimali klaida:	0.0760993193140892
: Faktinių (raudonų) ir modeliutų (mėlynų) grafinis palyginimas			

Experimentas nr. 2			
Įvesties parametrai (kiekis per dieną)		Esminis parametras (kiekis per dieną)	
1. Svetainėje pateikti IF pavedimai 2. Pasirašytos sutartys 3. Prisijungimų skaičius į svetainę 4. Nesėkmingų prisijungimų skaičius į svetainę 5. Užklausų skaičius į svetainę		->	Atidarytų puslapių skaičius svetainėje
Apmokymo laikas (s):	131.907	Apmokintą situacijų:	280 (iki 20071031)
Žinomų situacijų modeliavimo kokybė		Nežinomų situacijų modeliavimo kokybė	
SSE (sum squared error):	1.5854e-009/0	Nežinoma:	<u>48.44%</u>
Minimali klaida:	2.14642764256867e-008	Minimali klaida:	3.59538320163621e-005
Vidutinė klaida:	1.78574936211036e-006	Vidutinė klaida:	0.0206857129700733
Maksimali klaida:	1.08412974479527e-005	Maksimali klaida:	0.141267573740023
: Faktinių (raudonų) ir modeliuotų (mėlynų) reikšmių grafinis palyginimas			

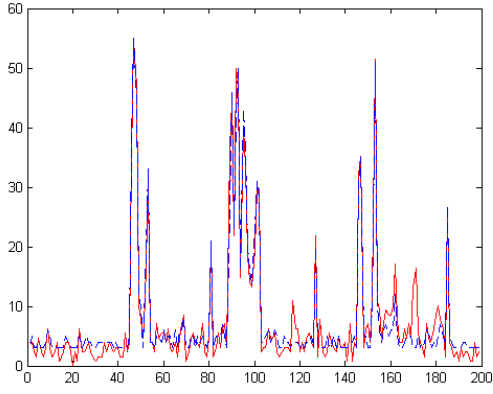
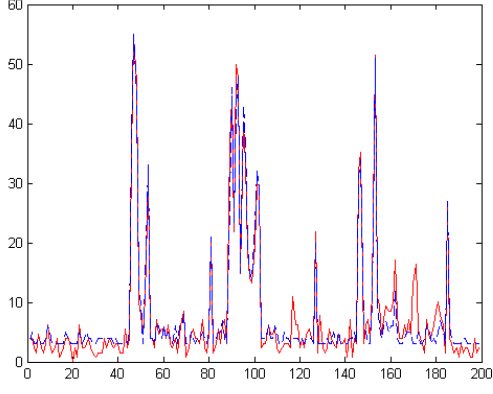
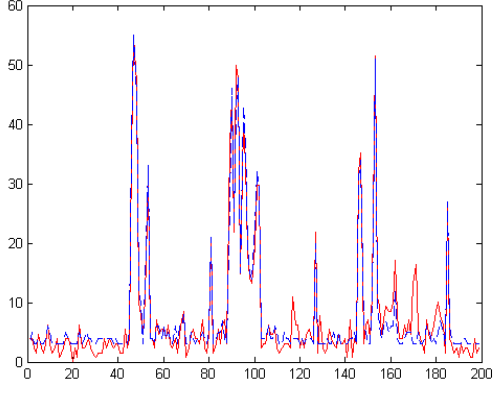


Experimentas nr. 3			
Įvesties parametrai (kiekis per dieną)		Esminis parametras (kiekis per dieną)	
1. Svetainėje pateikti IF pavidimai 2. Pasirašytos sutartys 3. VP pavidimai ofise 4. VP sandoriai biržoje 5. Prisijungimų skaičius į svetainę		->	Atidarytų puslapių skaičius svetainėje
Apmokymo laikas (s):	108.797	Apmokintą situacijų:	403 (iki 20071111)
Žinomų situacijų modeliavimo kokybė		Nežinomų situacijų modeliavimo kokybė	
SSE (sum squared error):	0.00242459/0	Nežinoma:	<u>25.78%</u>
Minimali klaida:	2.15121748024447e-006	Minimali klaida:	2.21835869104042e-006
Vidutinė klaida:	0.0017658554219935	Vidutinė klaida:	0.00791974778711906
Maksimali klaida:	0.00871092679492139	Maksimali klaida:	0.0551842286724512
: Faktinių (raudonų) ir modeliuotų (mėlynų) reikšmių grafinis palyginimas			



Experimentas nr. 4			
Įvesties parametrai (kiekis per dieną)		Esminis parametras (kiekis per dieną)	
1. Svetainėje pateikti IF pavedimai 2. Pasirašytos sutartys 3. VP pavedimai ofise 4. VP sandoriai biržoje 5. Prisijungimų skaičius į svetainę		->	Atidarytų puslapių skaičius svetainėje
Apmokymo laikas (s):	106.594	Apmokintą situacijų:	280 (iki 20071031)
Žinomų situacijų modeliavimo kokybė		Nežinomų situacijų modeliavimo kokybė	
SSE (sum squared error):	2.48919e-007/0	Nežinoma:	<u>48.44%</u>
Minimali klaida:	9.05011978552528e-009	Minimali klaida:	7.64657043586609e-005
Vidutinė klaida:	1.36631596762156e-005	Vidutinė klaida:	0.0288809812447356
Maksimali klaida:	0.000165418425251221	Maksimali klaida:	0.107130970613015
: Faktinių (raudonų) ir modeliuotų (mėlynų) reikšmių grafinis palyginimas			

Experimentas nr. 5			
Įvesties parametrai (kiekis per dieną)		Esminis parametras (kiekis per dieną)	
1. Svetainėje pateikti IF pavedimai 2. Pasirašytos sutartys 3. VP pavedimai ofise 4. VP sandoriai biržoje 5. Prisijungimų skaičius į svetainę		->	Atidarytų puslapių skaičius svetainėje
Apmokymo laikas (s):	118.953	Apmokintą situacijų:	492 (iki 20071117)
Žinomų situacijų modeliavimo kokybė		Nežinomų situacijų modeliavimo kokybė	
SSE (sum squared error):	4.78998e-005	Nežinoma:	<u>9.39%</u>
Minimali klaida:	2.2712298339278e-007	Minimali klaida:	4.00557752375796e-005
Vidutinė klaida:	0.000142630482188726	Vidutinė klaida:	0.0255311506170665
Maksimali klaida:	0.00391590248089588	Maksimali klaida:	0.798555680729683
: Faktinių (raudonų) ir modeliuotų (mėlynų) reikšmių grafinis palyginimas			

Faktinių (raudonų) ir modeliuotų (mėlynų) reikšmių grafinis palyginimas. Apmokyta iki 178 laiko vieneto (horizontali ašis), modeliuota po 178.	Neuroninio tinklo architektūra (neuronų skaičius sluoksniuose, atitinkamų sluoksnių perkelties funkcijos ir apmokymo metodas	Apsimokymo trukmė ir modeliavimo tikslumas (sum squared error)
	<pre>[15,7,15,1],{'tansig','tansig','tansig','purelin'},'trainbr'</pre>	76.656 s SSE 0.00314879526006
	<pre>[15,20,15,1],{'tansig','tansig','tansig','tansig','purelin'},'trainbr'</pre>	530.984 s SSE 0.00320357396622
	<pre>[10,5,10,1],{'tansig','tansig','tansig','tansig','purelin'},'trainbr'</pre> <p>(pakankamas variantas šiai parametų aibei)</p>	59.782 s SSE 0.00320299207326

3 lentelė. Trijų panašių neuroninių tinklų rezultatų palyginimas. Įvesties parametų aibė: dviejų procesų veikiančių OS-oje procesoriaus suvartojimas, o esminis parametras – bendras sistemos apkrovimas. Matome, kad šiai aibei užtenka mažiau neuronų turinčio tinklo – tikslumas pakankamas, o algoritmas veikia greičiau.

Išvados

Šiuo darbu pasiūlyta IT rizikų valdymo sistema, kuri geba modeliuoti rizikos parametrus, tokius kaip IS atsakymo laiką, IT įrangos apkrovimą, ryšių kanalų užimtumą ir kitus dydžius susijusius su IT rizikomis, pagal kitus parametrus, kurie yra stebimi susijusiose IS-ose. Sistema priima aibę istorinės statistinės informacijos iš IS-ų ir sugrupuoja ją pasitelkdama koreliacijų koeficientų metodą – taip sistema atranda ryšius, kurie iki tol galėjo būti ir nežinomi. Sistemai nurodoma, kuriuos statistinės informacijos parametrus norima modeliuoti. Tada kiekviena sudaryta grupė, kurioje yra bent vienas modeliuotinas parametras ir keli priežasties parametrai, yra perduodama pasiūlytos architektūros neuroniniam tinklui, kuris yra apmokomas ir taip iki tol nežinoma priklausomybė tarp tam tikro IT rizikos parametro ir pastarojo priežasties parametru yra apibendrinama. Šiam neuroniniam tinklui pateikus iki šiol nematytas priežasties parametru reikšmes galima modeliuoti kilsiančias rizikas. Sistema ypač naudinga esant poreikiui modeliuoti riziką, kuri yra sąlygota daugiau nei vieno parametro, kai priklausomybę žmogui yra per sudėtinga įvertinti.

Eksperimentai parodė, kad pasiūlyta neuroninio tinklo architektūra yra gana universali IT srities rizikų modeliavime, tačiau net ir šioje srityje skirtingom įvesties parametru aibėm galima atrasti konkrečius optimalesnius (greičio ir tikslumo prasme) neuroninius tinklus. Šis darbas rekomenduoja atlikti šią pirminę analizę, jei tik yra galimybė.

Kadangi sistema paremta neuroniniais tinklais, suprantama, jog jos rezultatų kokybė labai priklauso nuo pateikiamų apmokymui statistinių duomenų imties dydžio bei įvairovės. Dėl tos pačios priežasties modeliavimas tose informacijos parametru reikšmių srityse, kurios išeina iš jai pateiktų reikšmių rėžių, nėra korektiškas ir jau paliečia prognozavimo tematiką, žr. [MS97] ir [Gau97].

Literatūros sąrašas

- [PH04] Jeevan Perera, Jerry Holsomback.
An Integrated Risk Management Tool and Process, 2004.
- [SGF02] Gary Stoneburner, Alice Goguen, and Alexis Feringa.
Risk Management Guide for Information Technology Systems, 2002.
- [Mos95] Mr. Robin Moses.
Corporate Risk Analysis and Management Strategies, 1995.
- [JZ05] Xiaoxia Jia, Xizhao Zhou.
The Risk Identification and Monitoring of Electronic Business, 2005.
- [RL00] Janne Ropponen and Kalle Lyytinen.
Components of Software Development. How to Address Them? A Project Manager Survey, 2000.
- [SA04] Standards Australia.
HB 231:2004: Information security risk management guidelines, 2004.
- [CR99] Dhiman Chatterjee and VC Ramesh.
Real Options for Risk Management in Information Technology Projects, 1999.
- [EPR02] Peter J. Edwards, Andrew M. Peacock, David Renshaw, John M. Hannah, Alan F. Murray.
Minimizing Risk Using Prediction Uncertainty in Neural Network Estimation Fusion and Its Application to Papermaking, 2002.
- [Neu02] Donald E. Neumann.
An Enhanced Neural Network Technique for Software Risk Analysis, 2002.

- [YJZ06] Hu Yong, Chen Juhua, Rong Zhenbang, Mei Liu, Xie Kang.
A Neural Networks Approach for Software Risk Analysis, 2006.
- [NYD07] Meng Hai Ning, Qi Yong, Hou Di, Chen Ying.
Forecasting Software Aging of Service-Oriented Application Server Based on Wavelet Network with Adaptive Genetic Algorithm, 2007.
- [DP04] Brian Davis, Shirley Payne.
University of Virginia Information Technology Security Risk Management (ITS-RM) Program, 2004.
- [BM01] Andrew Blais, David Mertz.
An introduction to neural networks: pattern learning with the back-propagation algorithm, 2001.
<http://www.ibm.com/developerworks/library/l-neural/>
- [Die03] Dr. Frank Dieterle
Section 2.7. Neural Networks – Universal Calibration Tools, 2003.
http://drdieterle.com/phd/2_7.html
- [Sar01] Warren S. Sarle.
Neural-nets FAQ, Part 3 of 7: Generalization, 2001.
<http://faqs.org/faqs/ai-faq/neural-nets/part3/section-3.html>
- [Hea07] Heaton Research.
Chapter 5: Understanding Back Propagation, 2007.
<http://www.heatonresearch.com/articles/5/page2.html>
- [KY96] Tin-Yau Kwok, Dit-Yan Yeung.
Artificial Neural Networks - ICANN 96: Bayesian regularization in constructive neural networks, 1996.

- [DL20] Chi dung doan, shie-yui liong.
Generalization for multilayer neural network: bayesian regularization or early stopping, 200x.
- [DBH07] Howard Demuth, Mark Beale, Martin Hagan.
Neural Network Toolbox 5: User's Guide. Summary and Discussion of Early Stopping and Regularization, 2007.
- [Bou96] Paul Bourke.
Cross Correlation, Auto Correlation - 2D Pattern Identification, August 1996.
<http://local.wasp.uwa.edu.au/~pbourke/other/correlate/>
- [MS97] K. R. Muller, A. J. Smola, G.Ratsch, B. Scholkopf, J. Kohlmorgen, V. Vapnik.
Predicting Time Series with Support Vector Machines, 1997.
- [Gau97] Prediction with Gaussian Processes - From Linear Regression to Linear Prediction and Beyond, 1997.