

ПРОГНОЗИРАНЕ НА ПОСТЪПВАНЕТО И ОБСЛУЖВАНЕТО НА ПАКЕТИ В КОМУНИКАЦИИТЕ ЧРЕЗ РЕГРЕСИОНЕН АНАЛИЗ И НЕВРОННИ МРЕЖИ

Ивелина Балабанова

*Технически университет - Габрово, ул. „Хаджи Димитър“ № 4, Габрово, България
Кореспондиращ автор: ivstoeva@abv.bg*

SERVICE DEMAND FORECASTING IN COMMUNICATIONS BY REGRESSION ANALYSIS AND NEURAL NETWORKS

Ivelina Balabanova

*Technical University of Gabrovo, 4 Str. "Hadj Dimitar", Gabrovo, Bulgaria,
* Corresponding author: ivstoeva@abv.bg*

Abstract

The paper presents an approach developed for forecasting of traffic performance indices at the transmission medium monitoring procedures by classical regression analysis and artificial intelligence concept. The object of research are ICT structures with a queuing organization of the service of user requests. Adequacy assessments of Linear Regression and Cascade-Forward Neural Networks (CFNNs) were subsequently applied. Specific metrics to examine the quality have been analyzed, respectively Coefficient of determination R^2 , Mean-Squared Error - MSE, Mean Absolute Error - MAE and Root-Mean Square Error - RMSE. Significant levels of R^2 and required minimization of the errors defined during neural learning were obtained about the optimality and suitability requirements of the predictive models derived.

Keywords: transmission medium; service demand; predictive analysis; regression analysis; cascade-forward.

1. ВЪВЕДЕНИЕ

Регресионният анализ е един от най-често прилаганите аналитични способности в контекста на решаване на задачи за прогнозен анализ и оценка на значимостта и взаимовръзките между заложените променливи и обекти. Широкият спектър от изследвани области при интеграция на класическия анализ включва:

- бизнес, икономика, финанси, застраховане, маркетинг;
- земеделие, зоология, ветеринарна медицина и наука за климата;
- демография, здравеопазване;
- генетика, ксенобиология;
- недвижими имоти;
- транспорт, логистика [1-6].

Високата ефективност на изкуствени-те невронни мрежи е съизмерима с тази, постигана при класическите математи-

чески процедури при регресионно моделиране. Това се потвърждава от множеството сфери на тяхна интеграция при прогнозиране на вариациите на разнородни фактори, като по-съществените са:

- медицина - генетична диагностика, синтез на лекарства и медикаменти;
- метеорология – прогнозиране на валежите;
- фондова борса, финанси, електронна търговия, управление на риска;
- киберсигурност - детектиране на аномалии в мрежовия трафик, диагностика и оценка на заплахите, превенция срещу атаки [7-11].

В представеното проучване са изследвани възможностите за адаптиране на посочените статистически похвати в сферата на комуникациите. Докладът е насочен към разглеждане на един от ос-

новните аспекти при QoS в системи с опашкова организация на повикванията. В конкретният случай е разгледана имитационно моделирана комуникационна система с обект прогнозиране на количествени метрики на пакетното предаване на данни спрямо специфични индекси на преносната среда. Дефинирани са следните управляеми фактори:

- x_1 – Пропускателна способност (System Throughput - ST);
- x_2 – Сумарно време за престой в опашката и обслужване (System Response Time - SRT),
- и прогнозируем индекс на производителност:
- y – Комплексна метрика на постъпванията и обслужването на пакетна информация (System Demand - SD).

2. ИЗВЕЖДАНЕ НА АНАЛИТИЧНИ РЕГРЕСИОННИ МОДЕЛИ ЗА ПРОГНОЗИРАНЕ НА ИНДЕКС SD

На основата на линеен регресионен анализ бяха проведени последователни стъпки на извеждане и оценка на адекватността на аналитичен модел за прогнозиране на индекс на производителност SD спрямо входящи фактори ST и SRT. Първоначално беше извършена диагностика на формираната информационна извадка с включени ST, SRT и SD, представена на Фиг. 1.а), чрез способите на Дескриптивния анализ. Направена е оценка на статистическите показатели „Средна стойност Mean“ и „Стандартно отклонение Std. Dev.“ на дефинираните параметри, показана на Фиг. 1.б). Анализирана е тестова извадка със съдържание на 50 еталона за всеки индекс на преносната среда.

Проведен е регресионен анализ спрямо проверка на пригодността на модел (1), показан по-долу.

	1 x_1	2 x_2	3 y
1	43,666363	0,229009226	0,606
2	43,6344152	0,2291769	0,61837
3	43,6021152	0,229346672	0,63073
4	43,5694594	0,229518569	0,6431
5	43,5364445	0,22969262	0,65547
6	43,5030669	0,229868851	0,66784
7	43,4693232	0,23004729	0,6802
8	43,4352099	0,230227965	0,69257
9	43,4007237	0,230410904	0,70494
10	43,3658612	0,230596135	0,71731
11	43,330619	0,230783686	0,72967
12	43,2949935	0,230973588	0,74204
13	43,2589821	0,231165865	0,75441
14	43,2225813	0,231360546	0,76678
15	43,1857877	0,231557661	0,77914
16	43,1485984	0,231757238	0,79151
17	43,1110102	0,231959306	0,80388
18	43,07302	0,232163893	0,81624

а)

Variable	Means and Standard Deviations (QoS)		
	Means	Std.Dev.	N
x1	42,73106	0,605487	50
x2	0,23407	0,003326	50
y	0,90900	0,180284	50

б)

Фиг. 1. Извадка с трафични индекси на производителност а) и резултати от Дескриптивен анализ б)

$$y = b_0 + b_1x_1 + b_2x_2 \quad (1)$$

Резултатите от приложена линейна регресия са обобщени на Фиг. 2. Постигната е висока стойност на Коефициента на определеност $R^2 = 0.99994031$, показваща, че 99.994031 % от изменението на параметър „ y “ се дължи на въздействието на управляемите фактори „ x_1 “ и „ x_2 “. Всички изходни Коефициенти на регресия, както следва $b_0 = 144.905$, $b_1 = -1.835$ (отразяващ влиянието на x_1) и $b_2 = -280.170$ (свързан с оценка на въздействието на x_2), се определят като значими спрямо констатираната Вероятност „ p level“ много по-малка от прието Равнище на значимост $\alpha = 0.05$.

Regression Summary for Dependent Variable: y (QoS)						
R= ,99997016 R ² = ,99994031 Adjusted R ² = ,99993777						
F(2,47)=3937E2 p<0,0000 Std. Error of estimate: ,00142						
N=50	b*	Std. Err. of b*	b	Std. Err. of b	t(47)	p-value
Intercept			144,905	2,280890	63,5300	0,000000
x1	-6,16449	0,089758	-1,835	0,026725	-68,6791	0,000000
x2	-5,16703	0,089758	-280,107	4,865810	-57,5664	0,000000

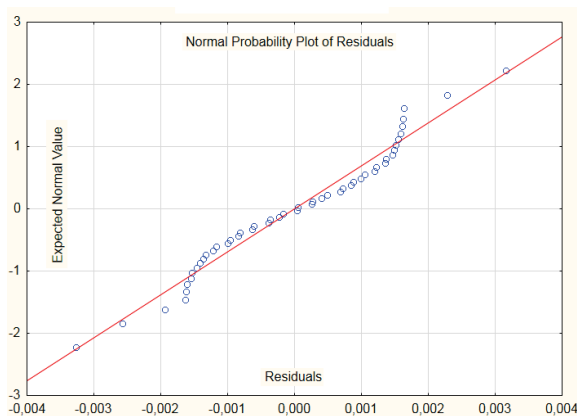
Фиг. 2. Резултати от линейна регресия

Критерият на Фишър $F(2,47) = 3937E2$ и съответната му вероятност $p < 0.0000$, отчетени спрямо $p \text{ level} \ll 0.05$, потвърждават висока адекватност на модел (1) при описание на входните данни. Крайният вид на синтезираният аналитичен модел за прогнозиране на вариациите на „Комплексна метрика на постъпванията и обслужването на пакети“ следва да бъде записан по-следният начин:

$$y = 144.905 - 1.835x_1 - 280.170x_2 \quad (1)$$

Case No.	Predicted & Residual Values (QoS)		
	Observed Value	Predicted Value	Residual
1	0.606000	0.609268	-0.003268
2	0.618370	0.620941	-0.002571
3	0.630730	0.632673	-0.001943
4	0.643100	0.644462	-0.001362
5	0.655470	0.656307	-0.000837
6	0.667840	0.668208	-0.000368
7	0.680200	0.680161	0.000039
8	0.692570	0.692167	0.000403
9	0.704940	0.704223	0.000717
10	0.717310	0.716328	0.000982
11	0.729670	0.728480	0.001190
12	0.742040	0.740677	0.001363
13	0.754410	0.752917	0.001493
14	0.766780	0.765198	0.001582
15	0.779140	0.777519	0.001622
16	0.791510	0.789876	0.001634
17	0.803880	0.802267	0.001613
18	0.816240	0.814691	0.001549

а)

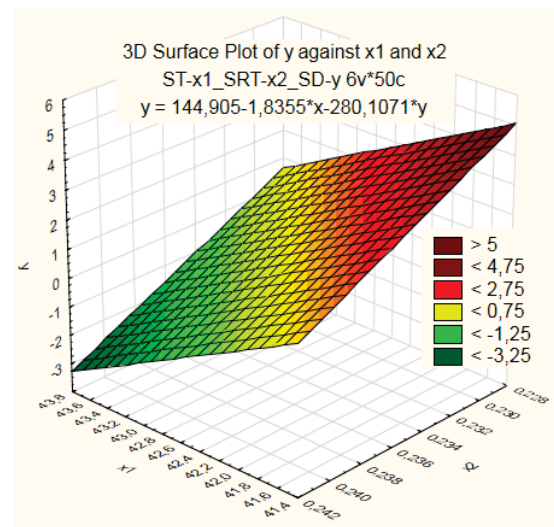


б)

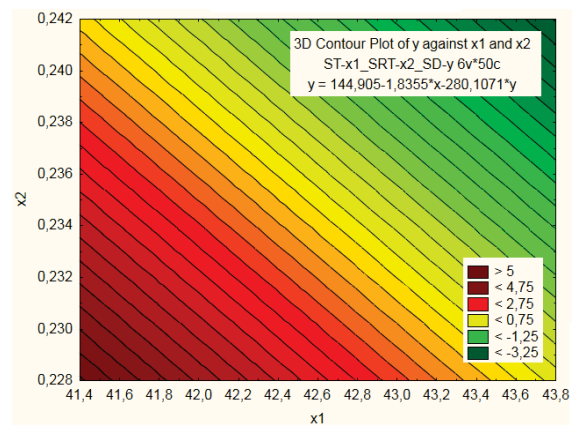
Фиг. 3. Теоретични и прогнозни стойности на индекс SD а) и Нормална вероятностна графика на Residuals б)

Във връзка с потвърждаване на ефективността на изведения аналитичен модел са представени количествените еквиваленти на наблюдаваните в хода на имитационно моделиране и получените прогнозни степени на индекса, категори-

зиращ постъпването и обработката на пакети на Фиг. 3.а). Допълнително е оценена коректността на приложения регресионен анализ чрез формиране на „Нормалната вероятностна диаграма на остатъците“ на Фиг. 3.б). Диаграмата дава индикация на Нормалното разпределение на Residuals, отчитащо се според близкото разположение на Остатъците по направление на линията от 45° - факт за коректност на анализа.



а)



б)

Фиг. 4. Повърхнина на отклика а) и Линии на еднакъв отклик $y = f(x_1, x_2)$ б)

Фигура 4 представя получената „Повърхнината на отклика“ $y = f(x_1, x_2)$ и полученото сечение на повърхнината с равнина, успоредна на равнината x_1Ox_2 , т.е. „Линиите на еднакъв отклик“ $y = f(x_1, x_2)$. Посочените зависимости на отклик потвърждават, че значителен дял от въз-

ходящите изменения на изходен параметър SD се получават при „средни и високи нива на фактор x_1 “ и „ниски и средни нива на фактор x_2 “.

3. ОБУЧЕНИЕ И ВЕРИФИКАЦИЯ НА CFNN ЗА ПРОГНОЗЕН АНАЛИЗ НА ИНДЕКС SD

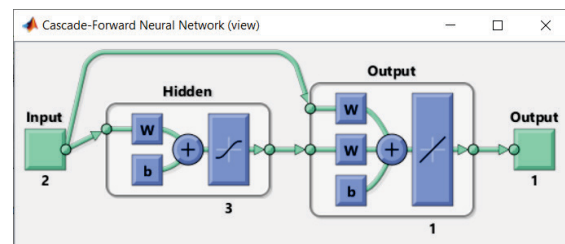
Изследвани са трислойни архитектури на Cascade-Forward Neural Networks в качеството на инструменти за прогнозен анализ. Заложено е обучение на моделите по Градиентен алгоритъм Levenberg-Marquardt и стъпково нарастване на невроните в скрития структурен слой. Оценени са въведени критерии MSE, MAE и RMSE при търсено изискване на минимизация като резултатите от анализа са поместени в Таблица 1. Търсеният оптимум - най-ниски степени $MSE = 2.3554 \cdot 10^{-10}$, $MAE = 1.1072 \cdot 10^{-5}$ и $RMSE = 1.5347 \cdot 10^{-5}$, беше намерен при идентично за указаните показатели начално количество от скрити неврони за приетия вариационен диапазон на анализ от 3 до 13 изчислителни единици. Максималните грешки от невронно обучение бяха установени при CFNN архитектура със съдържание на 7 междинни неврона за критерий MSE и 6 скрити неврона по отношение на метрики MAE и RMSE. Избраният краен CFNN модел за целите на прогнозната анализ е даден на Фиг. 5.

Табл 1. CFNNs за прогнозен анализ на индекс на производителност SD

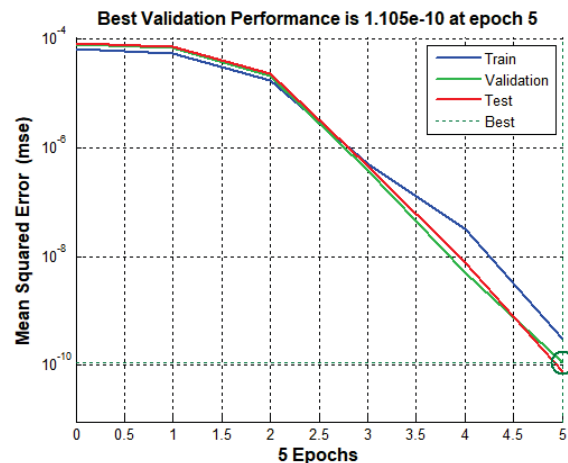
Скрити неврони	MSE	MAE	RMSE
3	2.3554e-10	1.1072e-05	1.5347e-05
4	8.2243e-10	2.4804e-05	2.8678e-05
5	3.8306e-10	1.4728e-05	1.9572e-05
6	5.3349e-10	1.9294e-05	2.3098e-05
7	9.8670e-10	2.3889e-05	3.1412e-05
8	7.0750e-10	6.4034e-05	8.4113e-05
9	2.8468e-10	3.6045e-05	5.3355e-05
10	5.5014e-10	3.2375e-05	7.4172e-05
11	7.1471e-10	1.2870e-05	2.6734e-05
12	8.8477e-10	1.2549e-05	2.9745e-05
13	4.6280e-10	3.6757e-05	6.8029e-05

Във връзка със селектираната архитектура на Фиг. 6 е представен хода на изменение на MSE за основните мрежови

процеси „обучение“, „валидиране“ и „тестване“. Наблюдавана е много близка тенденция на низходящо експоненциално изменение в поведението на кривите, описващи посоченият показател за оценка на качеството. Мрежовото обучение е проведено в обхвата на пет итерации, за които е достигната „най-добра валидираща производителност“ $1.105 \cdot 10^{-10}$. Според посочените факти не са открити индикации за „преобучение“ на анализирания невронна архитектура.

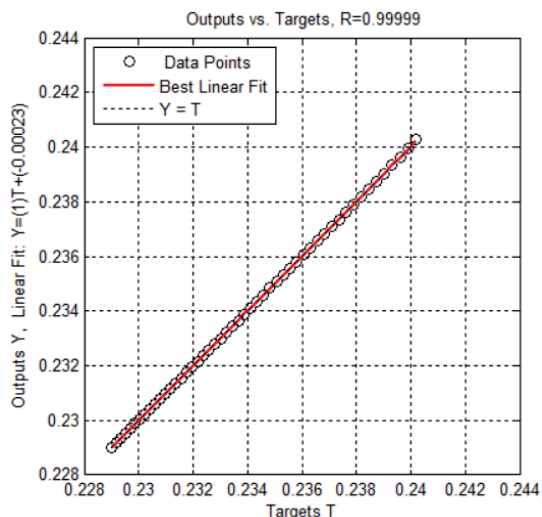


Фиг. 5. Селектиран CFNN модел за прогнозиране на индекс SD

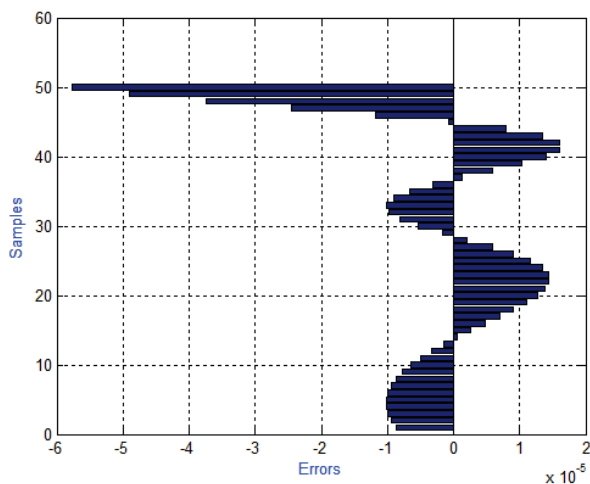


Фиг. 6. MSE при обучение, валидиране и тестване на CFNN за прогнозен анализ

Наблюдавано е почти идентично съвпадение между теоретичната и емпиричната линия на регресия за синтезирания CFNN модел на Фиг. 7. Регистрирана е значителна корелация $R = 0.99999$ между целевите и получените прогнозни стойности на SD с приложение на CFNN, потвърждаваща високата адекватност на синтезирания инструмент за регресионно моделиране.



Фиг. 7. Линии на регресия за синтезираната CFNN архитектура за прогнозиране на индекс на производителност SD



Фиг. 8. Диаграма на мрежовите грешки за избрания CFNN модел за прогнозен анализ

Фигура 8 представя положителните и отрицателните вариации на Остатъците, получени вследствие на прогнозните процеси с използване на CFNN (Фиг. 5). Установените разлики между Observed и Predicted стойности при оценка на „Комплексната метрика за пакетно предаване на информация“ се приемат за допустими, въпреки наличните отрицателни пикове на еталоните в края на обработваната информационна извадка.

4. ЗАКЛЮЧЕНИЕ

Висока степен на сходимост е установена между наблюдаваните в хода на

имитационно моделиране и получените прогнозни нива на метрика SD при линейна зависимост спрямо ST и SRT. Констатирано е преимущество на невронните мрежи в сравнение с регресионния анализ при наличие на експоненциална зависимост между управляемите фактори и прогнозирания индекс на производителност. Въпреки възникналото ограничение пред регресионния анализ се препоръчва продължаване на изследването при манипулации с други параметри на входящия и изходящия трафик в хода на търсене на полиномиални модели с висока адаптивност. По отношение на ефективността на Cascade-Forward Neural Networks се предвиждат процедури по сравнителен анализ спрямо допълнително въведени видове невронни мрежи при регресионна диагностика на ресурсната спецификация на преносната среда.

Благодарности: Изследванията в публикацията са реализирани и финансирани по Договор № 2406Е/ 2024г. „Мониторинг на мрежовия трафик и достъп на информация чрез биометрична диагностика с изкуствен интелект“.

ЛИТЕРАТУРА

- [1] Iqbal M. Application of regression techniques with their advantages and disadvantages. *Elektron Magazine* 2020;1(4):11-17.
- [2] Gaugler T., Stoeckl S., Rathgeber A. Global climate impacts of agriculture: a meta-regression analysis of food production. *Journal of Cleaner production* 2020;276(2020):1-10.
- [3] Rambotti S., Breiger R. Extreme and Inconsistent: a case-oriented regression analysis of health, inequality and poverty. *Socius Sage Journals* 2020;6(2020):1-8.
- [4] Alamin M., Sultana M., Xu H., Mollah M. Robustification of linear regression and its application in genome-wide association studies. *Frontiers Genetics Journal* 2020;11(549):1-17.
- [5] Banas M., Czaja J., Dabrowski J. The application of regression analysis for estimating

- the market value of commercial real estate. *Geomatics and Environmental Engineering Journal* 2022;16(1):17-35.
- [6] Rekha R., Abirami P., Aishvarya G., Akshaya B., Annapoorna A., Sanchana S. A web based application for tracking public transport and predicting usage. *International Journal of Aquatic Science* 2021;12(2):3770-3783.
- [7] Mansouri S. Application of neural networks in the medical field. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 2023;14(1):69-81.
- [8] Chima A., Duroha A. Artificial network application in prediction – a review. *African Journal of Computing & ICT* 2019;12(4):75-85.
- [9] Madhiarasan M., Louzazni M. Analysis of artificial neural networks: architecture, types and forecasting applications. *Journal of Electrical and Computer Engineering* 2022;2022(1):1-9.
- [10] González-Cortés D., Onieva E., Pastor I., Wu J. The application of artificial neural networks to forecast financial time series. *Logic Journal of the IGPL* 2024;2024(1):1-13.
- [11] Kharbanda V. Application of artificial intelligence in cyber security. *International Journal of Security and Privacy in Pervasive Computing* 2023;15(1):1-13.