

## NETWORK TRAFFIC MONITORING AN URBAN ENVIRONMENT BY MULTILAYER CFNNS

Ivelina Stefanova Balabanova

Technical University of Gabrovo  
Department of Communications Equipment and Technologies  
Gabrovo, Bulgaria  
ivstoeva@abv.bg

### Abstract

The paper examines one of the essential aspects of network traffic administration related to the analysis of the distribution of network traffic to corporate customers in urbanized areas in large city centers. An approach with the integration of four-layer Cascade-Forward Neural Networks (CFNNS) for regional traffic categorization is proposed. The approach includes the analysis and evaluation of the performance of several variants of multilayer neural structures with different ratios between the computational units in the hidden layers within set limits, respectively "1:2", "2:1" and "1:1". Selection procedures of multilayer CFNNS in search of an optimal ratio between achieving maximum accuracy and minimizing the Mean-Squared Error (MSE) and Mean Absolute Error (MAE) criteria were adopted. Diagnostics and verification of the qualities of the synthesized neural models using different categories of information samples were carried out.

**Keywords:** Network traffic; Traffic region identification; Cascade-Forward; Multilayer networks; Accuracy; MSE.

### ВЪВЕДЕНИЕ

Комплексният мониторинг на мрежовия трафик е една от основните дейности, регламентирани в политиките за достъп до информационни кълстери посредством WEB пространството на частни и корпоративни клиенти. По дефиниция се определя като многостранен процес за ефективно управление на трафичния поток, извършван от страна на операторите "доставчици на Интернет услуга", мрежовите администратори в учреждения и компании "клиенти" или различните регулаторни органи в сферата на информационното обслужване [1-3]. Основните направления в научните изследвания по отношение на анализ на мрежовия трафик обхваща различни процедури, които могат да се групират по следния начин:

- ✚ техники за кълстерен анализ: K – means, Fuzzy K-means, Auto class, DBSCAN,  $\gamma$ -Algorithm;
- ✚ подходи за класификация: Support Vector Machine (SVM) в комбинация с Rough Set Theory (RST), Principal Component Analysis (PCA);

Neural Networks с PCA, Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Bayesian Neural Networks (BNN), Deep Neural Networks (DNN);

- ✚ техники за многовариантен избор на решения Decision Tree: Random Forests;
- ✚ статистически подходи: Bayesian classification, Radial-Basis Function (RBF) networks, Multilayer Perceptron (MLP), C4.5 алгоритъм, Exploratory Factor Analysis (EFA) и др. [4-9].

В доклада е въведен подход за категоризация на мрежовия трафик на основата на:

- ✚ комплексен анализ на извадки с включени количествени параметри на пакетното предаване на данни – 1) Информационни потоци; 2) Доставяни пакети при IPv4, 3) Доставяни пакети при IPv6 и 4) Скорост на трансфер на данни;
- ✚ обучение, оценка и подбор на Cascade-Forward Neural Networks с използване на градиентен метод Levenberg-Marquardt,

във връзка с диагностика на доставяното Интернет съдържание към клиенти в урбанизирани градски райони.

### МНОГОСЛОЙНИ CASCADE-FORWARD NEURAL NETWORKS ЗА РАЗПОЗНАВАНЕ НА ГРАДСКИ ЗОНИ НА ПОТРЕБЛЕНИЕ НА ДОСТАВЯНО ИНТЕРНЕТ СЪДЪРЖАНИЕ

Проведени са изследвания, свързани с процедурен синтез на невронни инструменти и модели за идентификация на зони на потребление на доставяно пакетно съдържание посредством глобалната Интернет мрежа. Същността на процедурите се състои в оценка на приложимостта на Cascade-Forward Neural Networks. Приложени са невронни съотношения 1 – 2, 2 – 1 и 1 – 1 между невроните в скритите слоеве и дискретни кодови комбинации за асоцииране на принадлежността на трафичните параметри относно пакетното предаване към посочените класификационни групи. Въведена е последователност на функционалната активация на двата междинни и изходният слой, както следва бяха използвани “tansig”, “logsig” и “purelin” функции. Приети са индикатори за оценка на невронната производителност, респективно класификационна точност, средноквадратична грешка и средна абсолютна грешка.

*Таблица 1. Оценка на качеството на CFNNs за идентификация на зони на Интернет потребление при съотношение 1:2*

Скрит слой № 1	Скрит слой № 2	Точност, %	MSE	MAE
2	4	95.7	0.0455	0.0853
3	6	95.7	0.0343	0.1097
4	8	91.3	0.0668	0.1984
5	10	91.3	0.1221	0.1634
6	12	87.0	0.0700	0.1913
7	14	95.7	0.0281	0.0812
8	16	95.7	0.0489	0.1574
9	18	95.7	0.0570	0.1357
10	20	95.7	0.0474	0.1293
11	22	100.0	0.0288	0.1227
12	24	91.3	0.1263	0.2110
13	26	95.7	0.0660	0.1457
<b>14</b>	<b>28</b>	<b>100.0</b>	<b>0.0079</b>	<b>0.0669</b>
15	30	95.7	0.0574	0.0669

Таблица 1 обобщава отчетените нива на индикаторите за качеството при използвано базисно съотношение между невроните в двата скрити слоя на многослойните CFNNs. Тук беше установена минимална точност 91.3 % за модели с невронно съотношение „4:8“ и „5:10“. Най-висока точност 100.0 % беше постигната при CFNNs с включване на „11:22“ и „14:28“ междинни изчислителни единици. Най-високи индикатори MSE = 0.1263 и MAE = 0.2110 бяха регистрирани за многослойна архитектура при комбинация „12:24“. Минималните отчетени грешки, респективно MSE = 0.0079 и MAE = 0.0696, дават основание моделът със структурно съдържание на 14 и 28 неврона в двата скрити слоя на фиг. 1.а) да бъде избран в качеството на най-подходяща невронна архитектура за целева идентификация в рамките на разглеждания случай.

*Таблица 2. Оценка на качеството на CFNNs за идентификация на зони на Интернет потребление при съотношение 2:1*

Скрит слой № 1	Скрит слой № 2	Точност, %	MSE	MAE
4	2	95.7	0.0803	0.1655
<b>6</b>	<b>3</b>	<b>100.0</b>	<b>0.0052</b>	<b>0.0489</b>
8	4		0.0262	0.1233
10	5		0.0110	0.0707
12	6	95.7	0.0410	0.1577
14	7	87.0	0.1178	0.1863
16	8	100.0	0.0120	0.0887
18	9	87.0	0.0870	0.1567
20	10	91.3	0.0437	0.1039
22	11		0.0175	0.0945
24	12	100.0	0.0362	0.1500
26	13		0.0266	0.1178
28	14	95.7	0.0508	0.1614
30	15	100.0	0.0135	0.0861

По-ниски минимални MSE = 0.0052 и MAE = 0.0489 (таблица 2) бяха постигнати при второто базисно използвано съотношение 2 – 1 за комбинация с установена пълна коректност на разпознаване на еталоните от входящия набор „6:3“, показана на фиг. 1.б). Максимална точност

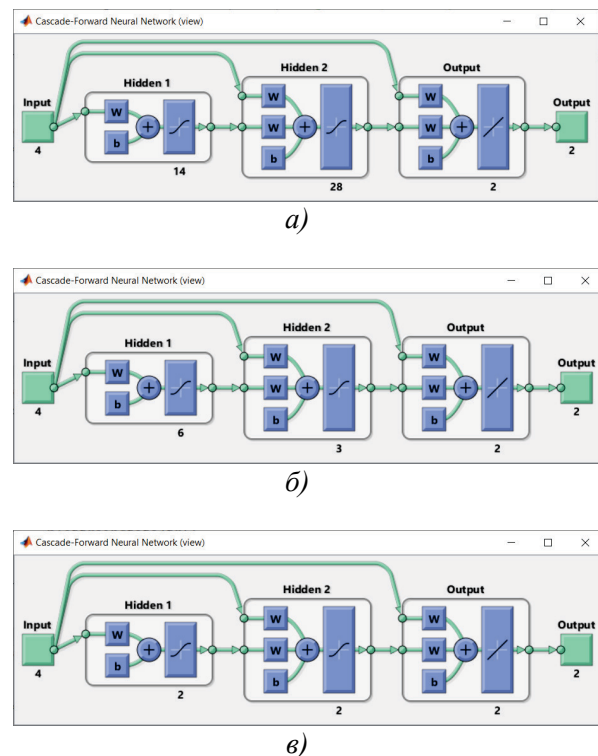
100.0 % е наблюдавана още при приложени невронни съотношения „8:4“, „10:5“, „22:11“, „24:12“ и „26:13“. Най-ниско показание на класификационната точност 87.0 % е отчетено за многослоен CFNN модел със съдържание на 14 в първи и 7 във втори междинен структурен слой.

**Таблица 3.** Оценка на качеството на CFNNs за идентификация на зони на Интернет потребление при съотношение 1:1

Скрит слой № 1	Скрит слой № 2	Точност, %	MSE	MAE
2	2	100.0	0.0101	0.0880
3	3	95.7	0.0210	0.0931
4	4	91.3	0.0651	0.1354
5	5	95.7	0.0451	0.1074
6	6	91.3	0.0756	0.2160
7	7	95.7	0.0356	0.1209
8	8	87.0	0.1025	0.1762
9	9	95.7	0.0248	0.0960
10	10	87.0	0.0902	0.2409
11	11	95.7	0.0651	0.1032
12	12	100.0	0.0135	0.0818
13	13	100.0	0.0241	0.1244
14	14	87.0	0.1308	0.2439
15	15	91.3	0.0667	0.1635
16	16	100.0	0.0189	0.0858
17	17	91.3	0.0650	0.1821
18	18	95.7	0.0623	0.1360
19	19	100.0	0.0374	0.1663
20	20	95.7	0.0513	0.1803

Според представените резултати в таблица 3 най-висока адекватност на CFNN инструментариума – еквивалент на точност 100.0 %, е присвоена на невронна структура при комбинация „2:2“ в рамките на изследванията от трета група със зададено базисно съотношение 1 – 1, дадена на фиг. 1.в). Тук регистрираните минимални MSE и MAE индикатори, равняващи се на 0.0101 (ниво над прага 0.0099) и 0.0880, определят посоченият CFNN модел с известни недостатъци в сравнение с изразените по-големи предимства на многослойният CFNN при базисно съотношение 2 – 1. Коректно разпознаване на приложените входни

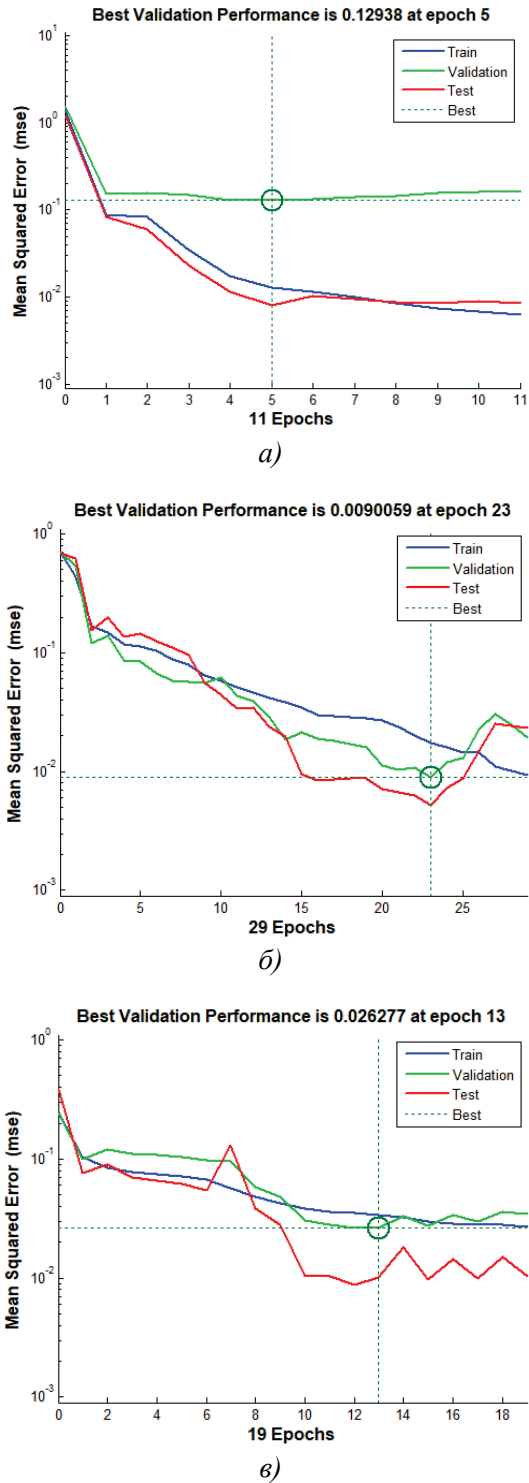
данни е проведено още при комбинации с идентични неврони в скритите слоеве „12:12“, „13:13“, „16:16“ и „19:19“. Максимални прагове на грешките MSE = 0.1308 и MAE = 0.2439 са регистрирани в хода на обучение на многослойна CFNN структура със съдържание на 14 неврона в двата междинни слоя.



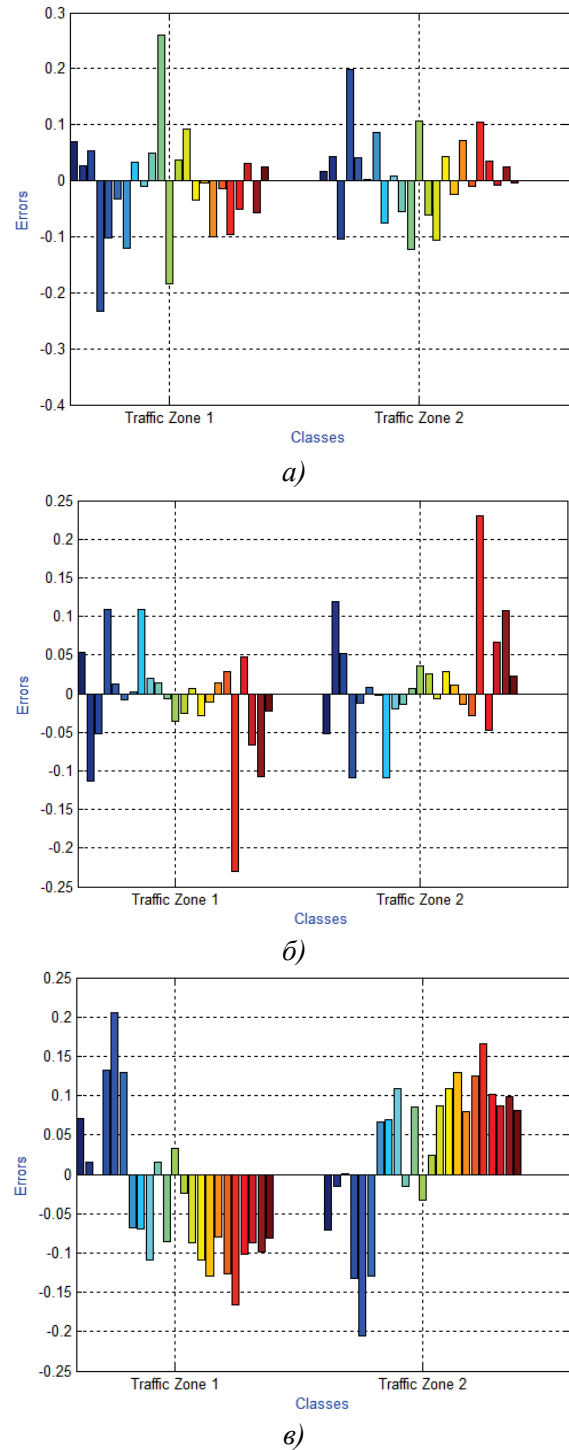
**Фиг. 1.** CFNNs за разпознаване на градски зони на разпределение на трафични пакети от глобалната Интернет мрежа при съотношения а) 1 – 2, б) 2 – 1 и в) 1 – 1

Съобразно диаграмите за анализ на изменението на средноквадратичната грешка от обучение, валидиране и тестване за синтезираните крайни CFNNs, генерирани на фиг. 2, е установена повторна тенденция на достигане на най-добра валидираща производителност преди преустановяване на обучаващите процедури. Налице е близък характер на изменение между „обучаващата и тестовата“, „валидиращата и тестовата“ и „обучаващата и валидиращата“ криви, описващи MSE, за последователно използваните съотношения 1 – 2, 2 – 1 и 1 – 1. Не е наблюдавано преобучение на многослойните невронни структури за разпознаване на обособени зони на тра-

фично Интернет потребление. Най-добра валидираща производителност, респективно 0.12938, 0.0090059 (потвърждаваща предимствана на съотношение 2 - 1) и 0.026277, са постигнати при 5-ти, 23-ти и 13-ти цикъл на обучение.



**Фиг. 2.** Диаграми на производителността за селектираните CFNNs за идентификация на градски зони на Интернет потребление при съотношения а) 1 – 2, б) 2 – 1 и в) 1 – 1



**Фиг. 3.** Диаграми на мрежовите грешки за избраните CFNNs за идентификация на зони на потребление на Интернет съдържание – а) 1 – 2, б) 2 – 1 и в) 1 – 1

Направена е оценка на мрежовите грешки чрез представените диаграми на фиг. 3. Спрямо трите изследвани базисни съотношения между изчислителните единици в скририте слоеве на CFNNs са констатирани вариации в допустимия

желан диапазон  $\pm 0.5$ . Регистрираните грешки попадат в следните граници:

- ✚ „-0.2323 до 0.2604“ за многослойната CFNN структура при съотношение между междинните неврони едно към две;
- ✚ „-0.2297 до 0.2297“ при CFNN със съотношение между скритите неврони две към едно;
- ✚ „-0.2048 до 0.2048“ за многослойната CFNN архитектура с използвано идентично базисно съотношение между скритите неврони.

## ЗАКЛЮЧЕНИЕ

Основавайки се на техниките на "дълбокото обучение", градиентните алгоритми и изборът на многослойни CFNN структури беше постигната висока ефективност спрямо реализираната задача за идентификация на региони на потребление на Интернет услуга от корпоративни клиенти. Разгледаният технически сценарий е близък до дейностите по управление на обслужването от големите мрежови оператори. Внедряването на технологията на Изкуственият интелект дават възможност за адаптивно и гъвкаво планиране на мрежовите ресурси по отношение на капацитет, надеждност и бързодействие при поддържане на високо QoS на услугата за конкретен регион или категория урбанизирани региони. Получените резултати изграждат сигурна база за разширяване на спектъра от аналитични инструменти с Машинно обучение. В подобни изследвания допълнително могат да бъдат засегнати аспекти за прогнозен анализ, оптимизация на специфични качествени и количествени

индекси на администрирания трафик, управление на сигурността чрез биометрична диагностика и др.

## REFERENCE

- [1] Kaur P. A Methodical review on network traffic monitoring and analysis tools. Journal of Composition Theory 2019; 12(9):1964-1968.
- [2] Mohammed A. Network Traffic Analysis: A Case Study of ABU Network. Intelligent Systems Engineering 2013; 4(4):33-40.
- [3] Zhou Y., Zhang D., Gao K., Sun Ch., Cao J., Wang Y., Xu M., Wu J. Intent-driven network traffic monitoring. IEEE/ACM Transactions on Networking 2021; 30(2):939-952.
- [4] Manish J., Hadi T. A review of network traffic analysis and prediction techniques. The Quarterly Journal 2015; 2015:1-22.
- [5] Lorez-Martin M., Carro B., Sanchez-Esguevillas A. Network traffic classifier with convolutional and recurrent neural networks for Internet of Things. IEEE Access 2017; 2017:1-9.
- [6] Michael A., Valla E., Neggatu N., Moore A. Network traffic classification via neural networks. University Cambridge 2017; 912:1-25.
- [7] Siracusano G., Galea S., Sanvito D., Haddadi H., Bifulco R. Re-architecting traffic analysis with neural network interface cards. In: the nineteenth symposium on networked systems design and implementation, vol. I, 2022, p. 513-533.
- [8] Ciptaningtyas H., Fatichah Ch., Sabila A. Network traffic anomaly prediction using artificial neural network, AIP Conf. Proc. 2017; 1818(1):1-15.
- [9] Alqudah N., Yaseen Q., Machine Learning for Traffic Analysis: A Review. Procedia Computer Science 2020; 170(2020): 911-916.