

Количествено идентифициране на телетрафични обслужващи устройства посредством дърво на решенията и ANN класификатор

Ивелина Балабанова, Георги Георгиев, Пенчо Пенчев

Quantitative identification of Teletraffic serving devices through a decision tree classifier and ANN: *There are classification models based on the decision tree and artificial neural network (ANN) to quantitatively identify the serving devices by teletraffic system $D + M + H_2 + E_3 / D / n / k$. Training classifiers were used: average downtime in the system W ; average waiting time in the queue W_q ; average number of calls in the system L and the average number of calls in queue L_q . An assessment of the model stood the method of decision tree by re-substitution and cross-validation at different levels of felling the branches. In the mean square error were studied changes and classification accuracy with a change of neurons in the hidden layer and the network structure is synthesized an artificial neural network. In the environment of the product is created LabVIEW virtual instrument for reading the data in Excel format and quantitatively identifying teletraffic serving devices.*

Key words: *Decision tree, Teletraffic system, Cross-validation, Artificial neural network, Classification models.*

ВЪВЕДЕНИЕ

Създадени са класификационни модели, базирани на дърво на решението и изкуствена невронна мрежа (ИНМ) за количествено идентифициране на обслужващи устройства на телетрафична система $D+M+H_2+E_3/D/n/k$. За обучение на класификаторите са използвани: средно време за престой в системата W ; средно време за чакане в опашката W_q ; среден брой повиквания в системата L и средния брой повиквания в опашката L_q . Извършена е оценка на модела, построен по метода дърво на решението, посредством ресубституция и крос-валидиране при различни нива на отсичане на разклонения. Изследвани са измененията на средноквадратичната грешка и класификационната точност с промяна на невроните в скрития мрежови слой и е синтезирана структура на изкуствена невронна мрежа. В среда на продукта LabVIEW е създаден виртуален инструмент за прочитане на данни в Excel формат и количествено идентифициране на телетрафични обслужващи устройства.

ИЗХОДНИ ДАННИ

Поставена е задачата за определяне на броя обслужващи устройства, при които са получени параметрите W , W_q , L и L_q на телетрафична система $D+M+H_2+E_3/D/n/k$ чрез имитационно моделиране. Приложени са два подхода за решение като са формирани класификационни модели по метода дърво на решението и апарата на изкуствените невронни мрежи. При обучение на класификаторите е използван набор от данни с независими предсказващи променливи:

- средно време за престой в системата W ;
- средно време за чакане в опашката W_q ;
- среден брой повиквания в системата L ;
- среден брой повиквания в опашката L_q .

Обособени са следните класификационни групи:

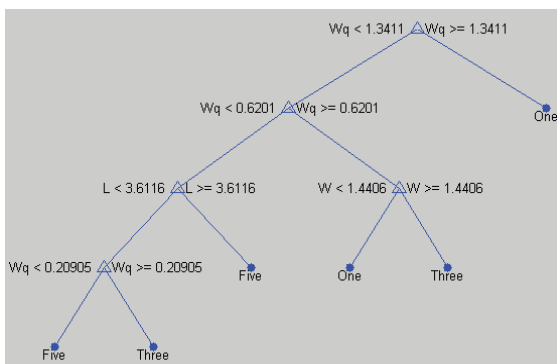
- клас 1 – едно обслужващо устройство;
- клас 2 – три обслужващи устройства;
- клас 3 – пет обслужващи устройства.

Входният набор от данни съдържа общо 57 наблюдения, съответно по 19 наблюдения в състава на всяка група. Кодирането на дефинирани класове се извършва посредством категориален (символен) тип данни при дърво на решението и дискретни кодови комбинации (0 и 1) и отдели изходни неврони при изкуствената

невронна мрежа. Позицията на единицата във всяка от комбинациите указва съответния брой обслужващи устройства.

КОЛИЧЕСТВЕНО ИДЕНТИФИЦИРАНЕ НА ТЕЛЕТРАФИЧНИ ОБСЛУЖВАЩИ УСТРОЙСТВА С ПРИЛОЖЕНИЕ НА ДЪРВО НА РЕШЕНИЕТО

Създаден е класификатор по метода дърво на решението, представляващ йерархична структура от логически "if-then" условия, за количествена идентификация на системните обслужващи устройства. На фиг.1 е представен първоначално генерирания класификационен модел.



Фиг. 1. Първоначално генериран модел по метода дърво на решението

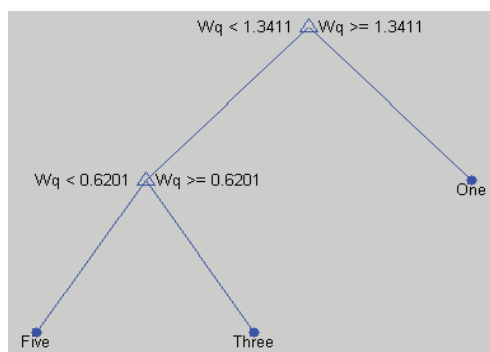
Приложени са методите ресубституция и крос-валидиране на оценка на точността при класификация. При оценяване на качествата на класификатора наблюденията от входния набор от данни са разпределени в две извадки - 75% за обучение и 25% за тестване на крос-валидиращите модели. Извършена е процедура по отсичане на разклонения от първоначално генерирания модел при различни нива с цел минимизиране на грешката от крос-валидиране и намиране на структура с оптимален брой възли. По правило точността при ресубституция се счита като по-оптимистична оценка, докато постигнатата при крос-валидиране се приема за по-достоверна при класифициране на данни, неучаствали в процеса на обучение. В табл. 1 са показани резултатите при оценка на качеството на класификационния модел.

Таблица 1. Резултати при оценка на качеството на класификатора

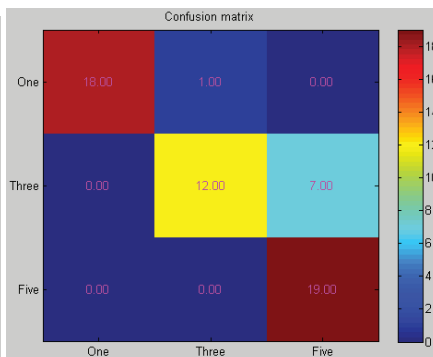
Ниво на отсичане	Точност при ресубституция, %	Точност при крос-валидиране, %	Приблизително очаквана точност при класифициране на нови данни, %
0	98.25	78.46	88.335
1	96.49	78.46	87.475
2	85.96	80.70	83.330
3	64.91	64.91	64.910
4	33.33	35.09	34.210

Установено е, че при второ ниво на отсичане на разклонения се получава най-висока класификационна точност от 80.70% при крос-валидиране. Въз основа на това за идентифициране на телетрафичните обслужващи устройства е избран класификационен модел, показан на фиг. 2. При съответното ниво на отсичане при ресубституция постигнатата точност е 85.96%. При класифициране на нови данни се приема приблизително очаквана точност от 83.33%.

Матрицата на коректни и некоректни класификации, представена на фиг. 3, показва коректно класифициране на наблюденията от трети клас. Едно наблюдение от първа и седем наблюдения от втора са некоректно отнесени, съответно към втора и трета класификационни групи.



Фиг. 2. Класификационен модел при второ ниво на отсичане



Фиг. 3. Класификационна матрица

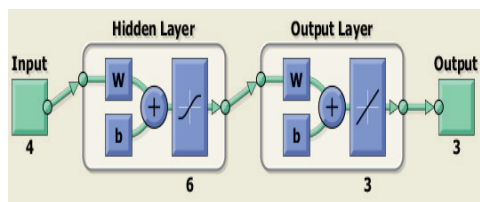
КОЛИЧЕСТВЕНО ИДЕНТИФИЦИРАНЕ НА ТЕЛЕТРАФИЧНИ ОБСЛУЖВАЩИ УСТРОЙСТВА С ПРИЛОЖЕНИЕ В ИЗКУСТВЕНА НЕВРОННА МРЕЖА

Таблица 2. Резултати при експериментиране с броя на скритите неврони

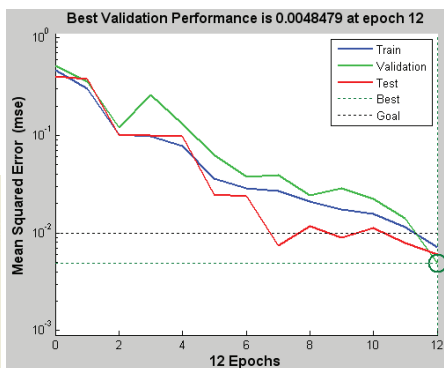
Скрити неврони	Точност, %	Средноквадратична грешка
5		0.0110
6		0.0060
7		0.0198
8	100.00	0.0246
9		0.0144
10		0.0351
11		0.0127
12		0.0901
13	88.90	0.0918
14		0.1360
15		0.0088
16	100.00	0.0084
17		0.0402
18	77.80	0.0768
19	100.00	0.0038
20	88.90	0.0204

Използвана е трислойна архитектура на изкуствена невронна мрежа (ИНМ) с обратно разпространение на грешката за идентифициране на телетрафични обслужващи устройства с тангес-сигмоидална и линейна функции на активация в скрития и изходен мрежови слоеве. Обучението на ИНМ е извършено по алгоритъма на Levenberg-Marquard при следните параметри epochs: 1000, show: 25, goal: 0.01, lr.: 0.05 и min_grad: 1-е^o5 и процентно съотношение на данните 70% за обучение, 15% за валидиране и 15% за тестване. Синтезирана е структура с шест скрити (междинни) неврона, при която са постигнати най-голяма и най-малка класификационна точност и средноквадратична грешка (mean squared error - MSE) от 100.00% и 0.0060. Резултатите при експериментиране с броя на невроните в междинния слой са показани в табл. 2.

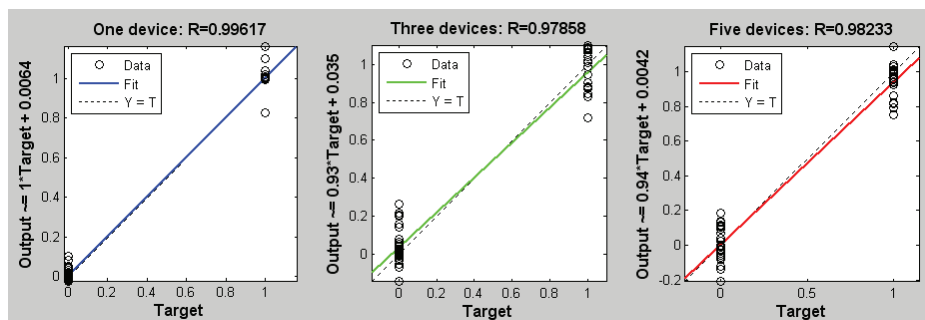
На фиг.4 и фиг.5 са представени избраната структура на ИНМ за количествено идентифициране и изменението на средноквадратичната грешка при обучение, валидиране и тестване.



Фиг. 4. Структура на ИНМ



Фиг. 5. Изменение на MSE



Фиг. 6. Линейни регресионни зависимости за изходните групи

Output Class \ Target Class	1	2	3	Total
1	2 22.2%	0 0.0%	0 0.0%	2 100%
2	0 0.0%	3 33.3%	0 0.0%	3 100%
3	0 0.0%	0 0.0%	4 44.4%	4 100%
Total	2 100%	3 100%	4 100%	9 100%

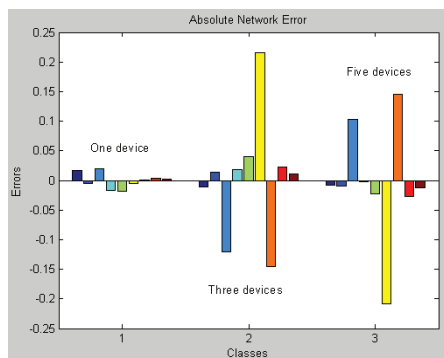
а)

Output Class \ Target Class	1	2	3	Total
1	19 33.3%	0 0.0%	0 0.0%	19 100%
2	0 0.0%	19 33.3%	0 0.0%	19 100%
3	0 0.0%	0 0.0%	19 33.3%	19 100%
Total	19 100%	19 100%	19 100%	57 100%

б)

Фиг. 7. Класификационни матрици за наблюдението а) от тестовата извадка и б) за целия набор от данни

Не са установени проблеми при обучение на невронната мрежа, валидиращата и тестова криви се характеризират с намаляващо изменение. Най-добра производителност при валидиране от 0.0048479 е достигната при 12-та итерация. Получените високи стойности на корелационните коефициенти R на линейните регресионни зависимости за изходните неврони, показани на фиг. 6, потвърждават много добра линейна връзка между мрежовите и целеви резултати. Наблюдава се добро групиране на данните около съответните логически нива. Наблюденията от тестовата извадка и целия входен набор от данни за класификационните матрици, представени на фиг. 7, са разположени по диагонал от горен ляв към долен десен ъгъл, а останалите матрични елементи са нули – свидетелство за пълно коректно определяне на груповата им принадлежност.



Фиг. 8. Абсолютна мрежова грешка

	1	2	3	4	5	6	7	8	9
1	1.0173	0.9951	0.0190	-0.0164	-0.0177	-0.0057	0.0012	0.0041	0.0023
2	-0.0111	0.0145	0.8801	1.0183	1.0402	0.2155	-0.1456	0.0232	0.0105
3	-0.0074	-0.0094	0.1030	-0.0018	-0.0225	0.7914	1.1459	0.9736	0.9880

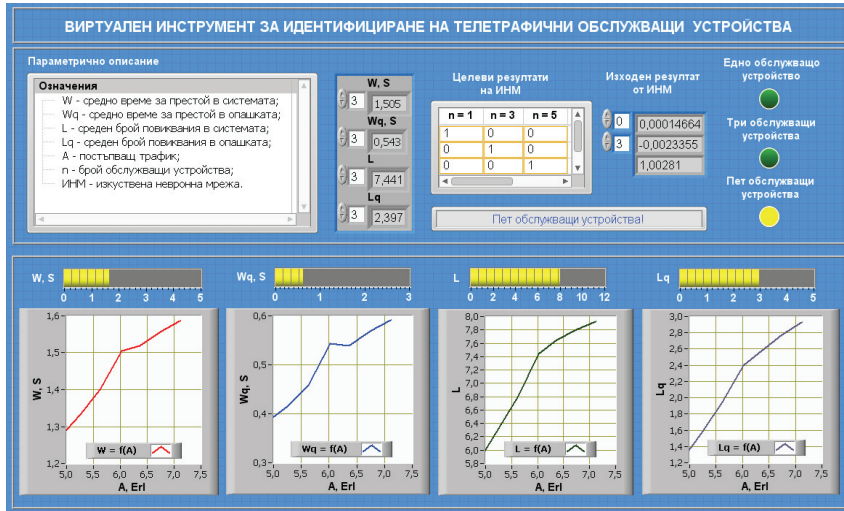
Фиг. 9. Мрежови резултати при класифициране на наблюденията от тестовата извадка

На фиг. 8 е показана абсолютната мрежова грешка, пресметната като разлика между получените резултати, представени на фиг. 9, и целеви резултати за ИНМ. Грешката се изменя в диапазона от -0.2086 до 0.2155 като най-ниски и най-високи нива на вариране се наблюдават, съответно за първа и трета класификационни групи.

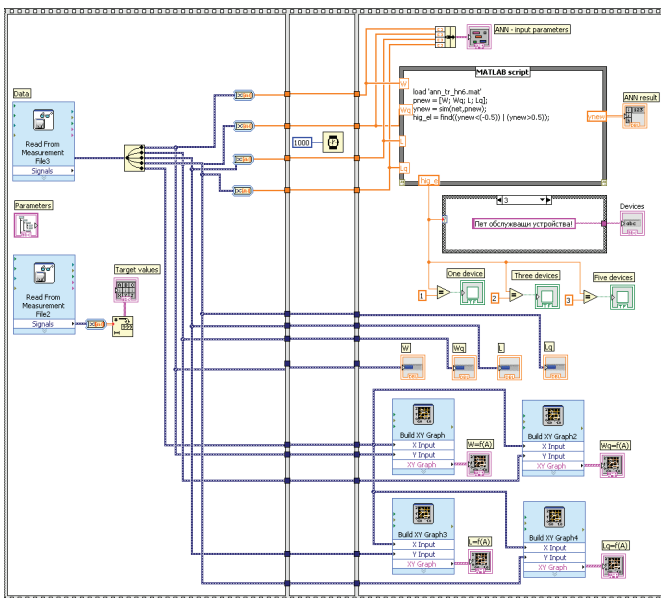
ВИРТУАЛЕН ИНСТРУМЕНТ ЗА КОЛИЧЕСТВЕНО ИДЕНТИФИЦИРАНЕ НА ТЕЛЕТРАФИЧНИ ОБСЛУЖВАЩИ УСТРОЙСТВА

В програмна среда на продукта LabVIEW е разработен виртуален инструмент (ВИ) за идентифициране на телетрафични обслужващи устройства. При тестване на работоспособността на ВИ са използвани данни, съхранявани в excel формат. След прочитане данните от тестовия файл, съответстващи на W ; W_q ; L и L_q , се конвертират от динамичен в скаларен тип и подават като входни променливи към matlab script. В скрипта се зареждат параметрите на обучената невронна мрежа за количествена идентификация. В калкулацията мрежови резултат се търси матричен елемент, чиято стойност е по-голяма от 0.5 или по-малка от -0.5. Спрямо позицията на елемента се извършва причисляване към съответна класификационна група. В

дървовидна структура е направено описание на използваните параметри. Посредством числови индикатори и индикатори на ниво се визуализират параметрите W ; W_q ; L и L_q и изходния мрежови резултат. В табличен вид са показани целевите резултати на ИНМ. Графични индикатори визуализират изменението на W ; W_q ; L и L_q спрямо постъпващия трафик в системата A . Идентифицираните системни обслужващи устройства се указват чрез светлинен индикатор и текстово съобщение. Предният панел и блоковата диаграма на ВИ са показани на фиг. 10.



а)



б)

Фиг. 10. Преден панел а) и блокова диаграма б) на ВИ

ЗАКЛЮЧЕНИЕ

Бурното развитие на мрежите с пакетна комутация, в частност Интернет и компютърните мрежи, доведе до появата на нов вид услуги, които стават все по-популярни сред потребителите на комуникационни услуги. Ако към това добавим и големия възход в областта на Интернет протоколите, които се превърнаха в предпочитан инструмент при пренос на информация за всички типове мрежи, поради отворения си интерфейс и широката реализация, то ние можем да кажем, че навлизаме в нова ера на комуникации. Технологиите за пренос на глас по Интернет протокол (VoIP), пренос на видео и видеоконференции по Интернет мрежата позволяват на потребителите да поддържат едно по-високо ниво на комуникативност и информированост. Успешно е илюстрирана възможността за приложение на използваните методи за класификация и виртуалния инструментариум LabVIEW в областта на съвременните телекомуникационни системи.

ЛИТЕРАТУРА

- [1] Радев Д., Т. Илиев, Г. Христов. Компютърно моделиране на телетрафични системи, Изд. РУ „Ангел Кънчев”, Русе, 2008.
- [2] MathWorks. Decision trees, 2015. <http://www.mathworks.com>.
- [3] Shih J., Hsu H., Multi-Channel Surface Acoustic Wave (SAW) Sensor Based on Artificial Back Propagation Neural (BPN) and Multivariable Linear Regression Analysis (MLR) for Organic Vapours. Journal of Chinese Chemical Society, 2007, 57, pp.401-410.
- [4] Artificial Intelligence: Foundation of Computation Agents. Neural networks, 2010. http://artint.info/html/ArtInt_183.html#nnet-ex.

За контакти:

гл. ас. д-р инж. Ивелина Балабанова, Катедра “Комуникационна техника и технологии”, Технически университет - Габрово”, тел: 066/827-375, e-mail: ivstoeva@abv.bg

ас. инж. Георги Георгиев, Катедра “Основи на електротехниката и електроенергетиката”, Технически университет - Габрово”, GSM: 0877522029, e-mail: givanow@abv.bg

доц. д-р инж. Пенчо Пенчев, Катедра “Комуникационна техника и технологии”, Технически университет - Габрово”, тел: 066/827-550, e-mail: ipkpen@tugab.bg

Докладът е рецензиран.

РУСЕНСКИ УНИВЕРСИТЕТ „АНГЕЛ КЪНЧЕВ“
UNIVERSITY OF RUSE „ANGEL KANCHEV“

ДИПЛОМА

Програмният комитет на
Научната конференция RU&SU'15
награждава с КРИСТАЛЕН ПРИЗ
“THE BEST PAPER”

**ИВЕЛИНА БАЛАБАНОВА,
ГЕОРГИ ГЕОРГИЕВ и ПЕНЧО ПЕНЧЕВ**
автори на доклада

“Количествено идентифициране на телетрафични
обслужващи устройства посредством дърво
на решенията и ANN класификатор”

DIPLOMA

The Programme Committee of
the Scientific Conference RU&SU'15
Awards the Crystal Prize
"THE BEST PAPER"

to **IVELINA BALABANOVA,
GEORGI GEORGIEV and PENCHO PENCHEV**
authors of the paper

“Identifying the parameters of Teletraffic model
of voice sources with application
of ANN and ANFIS classifiers”

РЕКТОР
RECTOR

проф. дтн Христо Белоев
Prof. DSc Hristo Beloev

10.10.2015