

РАЗПОЗНАВАНЕ НА ГАЗОВИ ЗАМЪРСИТЕЛИ НА ВЪЗДУШНАТА СРЕДА В ЗАТВОРЕНИ ПОМЕЩЕНИЯ

Георги Георгиев

ТУ-Габрово, гр. Габрово, ул. „Хаджи Димитър” 4, givanow@abv.bg

Резюме: В настоящата работа са представени резултати при разпознаване на летливи органични съединения (VOCs) с приложение на апарата на изкуствените невронни мрежи с обратно разпространение на грешката, както и с методите k – най-близки съседи и дърво на решенията. Синтезирани са трислойна структура на изкуствена невронна мрежа с тринадесет скрити неврона, класификатор с дефинирано метрично разстояние City block между най-близките пет k - съседи, както и класификационен модел при трето ниво на отсичане на разклонения по метода дърво на решенията.

Ключови думи: летливи органични съединения, изкуствена невронна мрежа, k – най-близки съседи, дърво на решенията.

1. Въведение

Идентифицирането на вредни и потенциално опасни за човешкото здраве газови замърсители на въздушната среда е основна предпоставка при осигуряване на здравословни и безопасни условия на труд, жизнен комфорт у дома, креативна и творческа атмосфера в училище и други. Процесът на газово разпознаване е задача, решавана от съвременните системи за мониторинг на качество на въздуха в затворени помещения [1-3].

От прилаганите методи при определяне на принадлежността на газови замърсители най-голям дял заема апаратът на изкуствените невронни мрежи [4-6]. Кодирването на присъствието и отсъствието на целеви газови съединения се извършва посредством променлива с различни нива, отделни неврони в изходния мрежови слой и кодови комбинации от -1 и +1 и други технически подходи. Сравнително по малко използвани методи в тази област са k – най-близки съседи и дърво на решенията [7, 8].

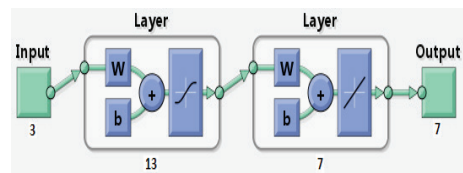
В настоящата работа е синтезирана структура на изкуствена невронна мрежа (ИНМ) и са селектирани класификатори за определяне на груповата принадлежност на газови замърсители, създадени по методите k – най-близки съседи и дърво на решенията.

2. Методи за разпознаване на VOCs

При обучение на класификаторите, създадени по указаните методи, са използвани експериментално получени данни за изходните съпротивления на газови сензори за следене на изменението на летливи органични съединения (VOCs), въглероден оксид (CO) и азотен диоксид (NO₂) при зададени концентрации на VOCs в тестова газова

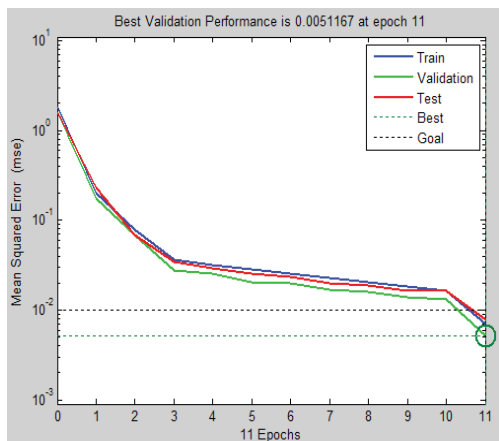
камера. По време на експеримента са анализирани проби от *Acetone (A)*, *Ethanol (E)*, *Trichloroethylene (TCE)*, *Dibutyl phthalate (D)*, *Xylene-o (X)*, *Benzene (B)* и *Toluene (TOL)*. На база на измерените сензорни съпротивления е формиран входен набор от данни, съдържащ общо 147 наблюдения - по 21 за всяка група.

Предложена е трислойна архитектура на ИНМ за разпознаване на VOCs. Дефинирането на газовото присъствие и отсъствие се извършва посредством комбинации от дискретни стойности (0 и 1). Позицията на единицата в дадена кодова комбинация указва принадлежността на съответна класификационна група. Входните данни са разделени в следните процентни съотношения, съответно 60% за обучение, 20% за валидиране и 20% за тестване на невронната мрежа, и нормализирани в интервала от -1 до +1. Мрежовото обучение е извършено по алгоритъма на Levenberg-Marquardt. Изследвани са измененията на средноквадратичната грешка и точността при класифициране с промяна на броя на невроните в скрития мрежови слой в диапазона от 5 до 20. Търсена е мрежова структура с минимална средноквадратична грешка и максимална достоверност при газово разпознаване, постигнати при 13 скрити неврона, съответно 0.0068 и 100.00%. На фиг.1 е представена структура на синтезираната ИНМ при разпознаване на VOCs.



Фиг.1. Структура на ИНМ

Изменението на средноквадратичната грешка при обучение, валидиране и тестване на невронната мрежа е показано на фиг.2. Процесът на мрежово обучение завършва при достигане на 11 итерация. Не се наблюдават индикации за преобучение на невронната мрежа, свидетелство за което е подобният и намаляващ характер на изменение на валидиращата и тестова криви.



Фиг.2. Средноквадратична грешка при обучение, валидиране и тестване на ИНМ

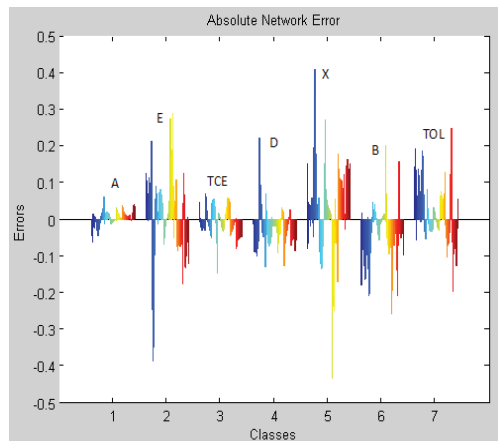
Достоверността от 100.00% съгласно получената класификационна матрица на фиг.3 потвърждава коректно класифициране на газовите съединения.

Confusion Matrix								
Output Class	1	2	3	4	5	6	7	
1	21 14.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
2	0 0.0%	21 14.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
3	0 0.0%	0 0.0%	21 14.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
4	0 0.0%	0 0.0%	0 0.0%	21 14.3%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
5	0 0.0%	0 0.0%	0 0.0%	0 0.0%	21 14.3%	0 0.0%	0 0.0%	100% 0.0%
6	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	21 14.3%	0 0.0%	100% 0.0%
7	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	21 14.3%	100% 0.0%
Target Class	1	2	3	4	5	6	7	100% 0.0%

Фиг.3. Класификационна матрица на ИНМ

На фиг.4 е показана абсолютната мрежова

грешка, пресметната като разлика между изходните и целеви резултати за ИНМ. По правило грешката не трябва да бъде по-голяма от 0.5, за да бъде класификацията коректна [9]. Получените нива на нейно изменение варират в границите от -0.4322 до 0.4106 , следователно условието за коректност е удовлетворено. Най-малки грешки се наблюдават при група А, а най-големи при клас Х.



Фиг.4. Абсолютна мрежова грешка

Създадени са модели за определяне на груповата принадлежност на VOCs посредством методите k -най-близки съседи и дърво на решенията. За оценка на качеството на класификаторите са използвани техническите подходи ресубституция и крос-валидиране. Класификационните групи са дефинирани посредством категориален тип данни като етикет за всеки клас е използвано съкращението на съответно съединение. Входният набор от данни е разделен на 75% за обучение и 25% за тестване на крос-валидиращите модели. Достоверността при ресубституция често се дава като по-оптимистична оценка, ето защо като по добра доверителна мярка се взема получената при крос-валидиране. При класифициране на нови данни се приема очаквана средна достоверност между двата подхода [10].

По метода k -най-близки съседи са създадени четири класификатора с различни метрични единици за разстояния между най-близките съседи, съответно:

- o Евклидово разстояние;
- o разстояние на Минковски;
- o разстояние City block;
- o разстояние на Чебишев.

За всеки класификационен модел са изслед-

вани измененията на получените некоректни класификации и класификационни точности с промяна на параметъра k в границите от 1 до 10. В табл. 1 са представени получените резултати за селектирания най-добър класификатор с дефинирано разстояние City block.

При оценка на качеството на класификатора посредством ресубституция, най-висока достоверност 100.00% е постигната при k=1, а най-ниска - при стойности на параметъра k=9 и k=10, съответно 80.95%. При k=5 е получена достоверност 85.03%. По метода крос-валидиране са постигнати най-висока и най-ниска достоверности 80.95% и 76.19%, съответно при пет и девет най-близки съседа. За определяне на груповата принадлежност на органичните съединения е избрана стойност на параметъра k=5. При класифициране на данни, които не са

участвали при обучение и тестване на модела, се очаква достоверност 82.99%.

Получената матрица на коректни и некоректни класификации, представена на фиг.5, показва 100.00% коректно класифициране на наблюдаваната от първа, четвърта и шеста класификационни групи.

Класификатор, построен по *метода дърво на решенията* за определяне на принадлежността на целевите газови съединения е показан на фиг.6. В табл. 2 са показани получените резултати при оценка на качеството на класификационните модели посредством указаните технически подходи при различни нива на отсичане на разклонения от първоначално генерирания модел.

При приложение на ресубституция е установена тенденция за намаляване на достоверността с увеличаване на премахнатите възли в

Таблица 1. Класификатор с дефинирано метрично разстояние City block

k – най-близки съсед	Ресубституция		Крос-валидиране		Очаквани параметри при класифициране на нови данни	
	Некоректни класификации, %	Дост., %	Некоректни класификации, %	Дост., %	Некоректни класификации, %	Дост., %
1	0.00	100.00	22.45	77.55	11.225	88.775
2	14.97	85.03	23.13	76.87	19.050	80.950
3	14.29	85.71	22.45	77.55	18.370	81.630
4	15.65	84.35	22.45	77.55	19.050	80.950
5	14.97	85.03	19.05	80.95	17.010	82.990
6	15.65	84.35	23.81	76.19	19.730	80.270
7	17.01	82.99	21.77	78.23	19.390	80.610
8	17.01	82.99	21.77	78.23	19.390	80.610
9	19.05	80.95	23.81	76.19	21.430	78.570
10	19.05	80.95	22.45	77.55	20.750	79.250

Таблица 2. Класификационни модели, построени по метода дърво на решенията

Ниво на отсичане	Ресубституция		Крос-валидиране		Очаквани параметри при класифициране на нови данни	
	Некоректни класификации, %	Дост., %	Некоректни класификации, %	Дост., %	Некоректни класификации, %	Дост., %
0	7.48	92.52	23.13	76.87	15.305	84.695
1	8.16	91.84	23.81	76.19	15.985	84.015
2	12.24	87.76	23.81	76.19	18.025	81.975
3	20.41	79.59	23.81	76.19	22.110	77.890
4	30.61	69.39	33.33	66.67	31.970	68.030
5	43.54	56.46	45.58	54.42	44.560	55.440
6	71.43	28.57	53.06	46.94	24.490	75.510
7	85.71	14.29	85.71	14.29	85.710	14.290

структурата на дървото. Най-висока и най-ниска достоверности 92.52% и 14.29% се наблюдават при нулево и седмо ниво на отсичане на разклонения. При крос-валидиране най-висока достоверност 76.87% е постигната при нулево ниво, а най-ниска - при последно ниво на отсичане на разклонения, съответно 14.29%. Целта на процедурата по отсичане на разклонения се състои в намирането на ниво, минимизиращо получената част на некоректни класификации при крос-валидиране и гарантиращо структура с оптимален брой възли. Базирайки се на това, за идентифициране на газовите замърсители е избран класификационен модел при трето ниво на отсичане на разклонения. При оценка на качеството на избрания класификатор са получени достоверности от 79.59% и 76.19%, съответно при ресубституция и крос-валидиране. Нови наблюдения ще бъдат класифицирани с очаквана достоверност 77.89%.

На фиг.7 е представена получената класификационна матрица. Наблюдава се 100.00% коректно класифициране на наблюденията за класове А и В.

3. Заключение

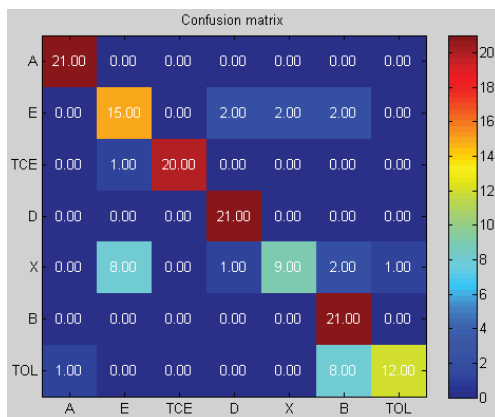
Посредством синтезираната структура на изкуствена невронна мрежа е постигнато 100.00% коректно разпознаване на VOCs. Очакваната достоверност при класифициране на нови данни с приложение на моделите, построени по методите k – най-близки съсед и дърво на решенията, е съответно 82.99% и 77.89%. Представените резултати показват, че при определяне на груповата принадлежност на тестовите съединения най-голям процент коректно разпознаване се получава при приложение на апарата на изкуствените невронни мрежи, последван от методите k – най-близки съсед и дърво на решенията.

4. Литература

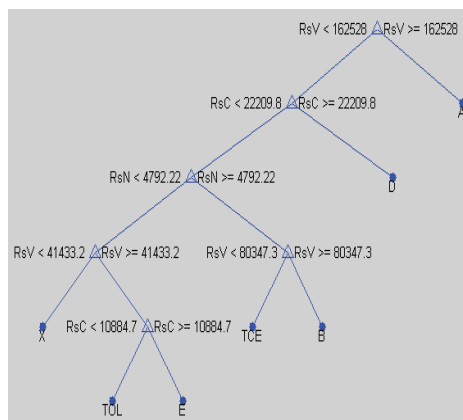
[1] World Health Organization – Regional Office for Europe. WHO Guidelines for Indoor Air Quality: Selected Pollutants, 2010, pp.1-484.

[2] M. Freitas, et al.*. Indoor Air Quality in Primary Schools, Environmental Health and Air Pollution Case Studies, DOI: 105772/17609, 2011, pp. 361-384.

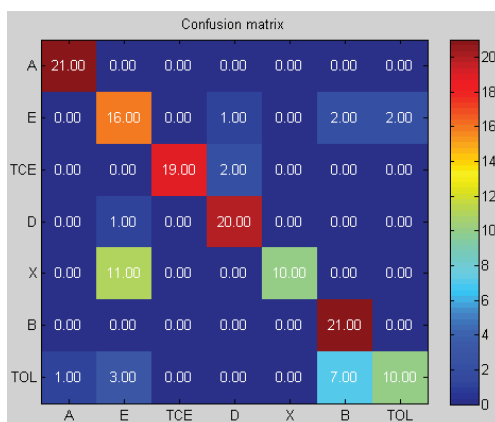
[3] California Environmental Protection Agency. Ventilation and Indoor Air Quality in New Homes, California Energy Commission, PIER Collaborative Report, CEC-500-2009-085, 2009, pp.1-426.



Фиг.5. Класификационна матрица при k=5, получена при крос-валидиране



Фиг.6. Класификационен модел при ниво на отсичане на разклонения 3



Фиг.7. Класификационна матрица, получена на база на избрания модел

[4] **J. Shih, H. Hsu.** Multi-Channel Surface Acoustic Wave (SAW) Sensor Based on Artificial Back Propagation Neural (BPN) Network and Multivariable Linear Regression Analysis (MLR) for Organic Vapours. Journal of Chinese Chemical Society, 54, 2007, pp.401-410.

[5] **Z. Nenova, G. Dimchev.** Compensation of the Impact of Disturbing Factors on Gas Sensor Characteristics, Acta Polytechnica Hungarica, Vol.10, No 3, 2013, pp.97-111.

[6] **З. Ненова, Г. Димчев.** Разпознаване на газове с помощта на изкуствени невронни мрежи. Известия на ТУ – Габрово, No 45, 2013, с.45-50.

[7] **S. Belhaouari, R. Abaza,** Gas Identification by Using of a Cluster-K-Nearest Neighbor, International Conference of Machine Learning and Computing, Vol.3, 2011, pp.1123-127.

[8] **M. Hassan, A. Bermak.** Gas Classification

Using Binary Decision Tree Classifier, Circuit and Systems (ISCAS), IEEE International Symposium, 2014, pp.2579-2582.

[9] **S. Haykin.** Neural Networks: A Comprehensive Foundation, Pearson Prentice Hall Inc., 2005, pp.1-823.

[10] <http://www.mathworks.com/help/stats/classification-using-nearest-neighbors.html>.

Данни за автора:

Георги Иванов Георгиев, Маг. инж. специалност „Безопасност на труда“ (2008г.), Асистент (2009г.), катедра „Основи на електротехниката и електроенергетиката“, факултет “Електротехника и електроника”, ТУ-Габрово. Научни интереси: измервания и контрол на параметри на въздушната среда, бази от данни и др.

RECOGNITION OF INDOOR GASEOUS AIR POLLUTIONS

Georgi Georgiev

TU-Gabrovo, Gabrovo, 4 “H. Dimitar”, givanow@abv.bg

Abstract: In this paper are presented the results at recognition of volatile organic compounds (VOCs) by back-propagation neural networks and methods of k - nearest neighbors and decision tree. The three-layer architecture of an artificial neural network with thirteen hidden neurons, a classifier with defined city block metric distance between five k – nearest neighbors and classification model at third pruning level of tree are selected.

Key-Words: volatile organic compounds, artificial neural network, k – nearest neighbors, decision tree.

References

[1] World Health Organization – Regional Office for Europe. WHO Guidelines for Indoor Air Quality: Selected Pollutions, 2010, pp.1-484.

[2] **M. Freitas, et al.*.** Indoor Air Quality in Primary Schools, Environmental Health and Air Pollution Case Studies, DOI: 105772/17609, 2011, pp. 361-384.

[3] California Environmental Protection Agency. Ventilation and Indoor Air Quality in New Homes, California Energy Commission, PIER Collaborative Report, CEC-500-2009-085, 2009, pp.1-426.

[4] **J. Shih, H. Hsu.** Multi-Channel Surface Acoustic Wave (SAW) Sensor Based on Artificial Back Propagation Neural (BPN) Network and Multivariable Linear Regression Analysis (MLR) for Organic Vapours. Journal of Chinese Chemical Society, 54, 2007, pp.401-410.

[5] **Z. Nenova, G. Dimchev.** Compensation

of the Impact of Disturbing Factors on Gas Sensor Characteristics, Acta Polytechnica Hungarica, Vol.10, No 3, 2013, pp.97-111.

[6] **Z. Nenova, G. Dimchev.** Razpoznavane na gazove s pomoshta na izkustveni nevronni mrezhi. Izvestiya na TU – Gabrovo, No 45, 2013, s.45-50.

[7] **S. Belhaouari, R. Abaza,** Gas Identification by Using of a Cluster-K-Nearest Neighbor, International Conference of Machine Learning and Computing, Vol.3, 2011, pp.1123-127.

[8] **M. Hassan, A. Bermak.** Gas Classification Using Binary Decision Tree Classifier, Circuit and Systems (ISCAS), IEEE International Symposium, 2014, pp.2579-2582.

[9] **S. Haykin.** Neural Networks: A Comprehensive Foundation, Pearson Prentice Hall Inc., 2005, pp.1-823.

[10] <http://www.mathworks.com/help/stats/classification-using-nearest-neighbors.html>.

РАСПОЗНАВАНИЕ ГАЗОВЫХ ЗАГРЯЗНЕНИЙ ВОЗДУХА В ЗАКРЫТЫХ ПОМЕЩЕНИЯХ

Георги Георгиев

ТУ-Габрово, Габрово, „Хаджи Димитър” 4, givanow@abv.bg

Резюме: В работе представлены результаты обнаружения летучих органических соединений (VOCs) с использованием аппарата искусственных нейронных сетей с обратным распространением ошибки, как и методов k-ближайших соседей и дерева решений. Синтезированы трехслойная структура искусственной нейронной сети с тринадцатью скрытыми нейронами, классификатор с определенным расстоянием метрики City block между пятью ближайшими k-соседями и модель классификации при третьем уровне отсекания разветвлении по методу дерева решений.

Ключевые слова: летучие органические соединения, искусственная нейронная сеть, k - ближайшие соседи, дерево решений.