

# ПРИМЕНЕНИЕ «DEEP LEARNING» ПРИ ПРОГНОЗИРОВАНИИ ПОЖАРООПАСНЫХ ПОКАЗАТЕЛЕЙ КИСЛОРОДСОДЕРЖАЩИХ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ

**Д.С. Королев;**

**А.В. Калач, доктор химических наук, профессор.**

**Воронежский институт ГПС МЧС России**

В целях уменьшения погрешности при определении пожароопасных свойств органических веществ при помощи способа прогнозирования на основе молекулярных дескрипторов и искусственных нейронных сетей, реализуемых оригинальной компьютерной программой «Нейропакет КДС 1.0», было решено модифицировать имеющийся программный продукт путем внедрения компьютерной библиотеки Deep Learn Toolbox. Представленная библиотека реализует моделирование искусственных нейронных сетей путем «глубокого обучения». В этом случае искусственная нейронная сеть имеет два и более скрытых слоя. Проводили верификацию данных, основываясь на некоторых справочных данных. Установлено, что полученные данные в результате работы модифицированной программы дают относительную погрешность по сравнению со справочными данными, не превышающую 5 %.

*Ключевые слова:* глубокое обучение, молекулярные дескрипторы, пожарная безопасность, свойства веществ, искусственные нейронные сети

## APPLICATION OF «DEEP LEARNING» WHEN FORECASTING FIRE-DANGEROUS INDICATORS OF OXYGEN-CONTAINING ORGANIC COMPOUNDS

D.S. Korolev; A.V. Kalach. Voronezh institute of State fire service of EMERCOM of Russia

For the purpose of reduction of an error in case of determination of fire-dangerous properties of organic substances by means of a forecasting method on the basis of molecular descriptors and artificial neural networks, realized by the original computer program «Neyropaket KDS 1.0» it was decided to modify the available software product by implementation of computer library Deep Learn Toolbox. The provided library realizes modeling of artificial neural networks by «deep training». In this case the artificial neural network has two and more hidden layers. Carried out verification of data, based on some help data. It is established that the obtained data as a result of work of the modified program give a relative error in comparison with help data not exceeding 5 %.

*Keywords:* deep training, molecular descriptors, fire safety, properties of substances, artificial neural networks

В настоящее время известно более 100 млн органических веществ, а их пожароопасные свойства изучены частично и лишь для нескольких тысяч соединений. Это связано в первую очередь с тем, что материальные и временные затраты на экспериментальное определение полного набора характеристик пожароопасных свойств для каждого вещества колоссальны. Быстро решить эту задачу практически невозможно, поэтому расчетные методы, компьютерное моделирование могут рассматриваться как альтернатива экспериментальному подходу [1]. Поэтому неудивительно, что в различных

научных журналах появляются статьи с новыми уравнениями и способами для прогнозирования пожароопасных показателей.

Таким образом, создание нового конкурентоспособного способа прогнозирования пожароопасных показателей различных органических соединений является актуальным.

Следует отметить, что впервые в работах [2, 3] авторами использовался способ прогнозирования пожароопасных свойств веществ на базе молекулярных дескрипторов и искусственных нейронных сетей, который реализуется при помощи разработанной оригинальной компьютерной программы «Нейропакет КДС 1.0» [4].

Работоспособность программы основана на проектировании персептронов, которые адекватно реагируют на предоставленные примеры (молекулярные дескрипторы изученных веществ). Причем при увеличении числа нейронов внутреннего слоя персептрона погрешность обучения обычно падает. Таким образом, моделировалась искусственная нейронная сеть с «N» – количеством входов, одним выходом и одним скрытым слоем [5], но современные компьютерные технологии не стоят на месте и в 2006 г. в мире сделали открытие – так называемое обучение глубоких искусственных нейронных сетей. Поэтому все чаще можно встретиться с таким термином, как «deep learning» (глубокое обучение) [6]. Этот новомодный термин повествует о том, как собрать из каких-то простых абстракций более сложную и глубокую абстракцию (репрезентацию), притом, что даже самые простые абстракции должен собирать сам компьютер, а не человек. То есть речь уже не просто об обучении, а о метаобучении, при котором можно эффективно предобучать многослойную искусственную нейронную сеть, а именно: обучать каждый слой отдельно при помощи ограниченной машины Больцмана, а затем дообучать при помощи метода обратного распространения ошибки [7].

Образно говоря, компьютер самостоятельно должен научиться, как лучше ему учиться. И, по сути, термин «глубокое» применяется к искусственным нейронным сетям, где используется больше одного скрытого слоя, поэтому формально «глубокий» значит ещё и более глубокая архитектура нейронной сети.

Уникальным для глубокого обучения является то, что машина сама находит признаки и структурирует их иерархично: из более простых складываются более сложные. На рис. 1 представлен пример работы искусственной нейронной сети с традиционным и глубоким обучением соответственно.

Анализируя рис. 1, видно, что при обычном обучении имеется всего один скрытый слой, при котором процесс обучения происходит хаотично. Полная противоположность ему – «глубокое обучение» искусственной нейронной сети. В этом случае имеется два и более скрытых слоев, содержащих полезную информацию, при которых программа начинает на первом уровне извлекать самые базовые, структурно простые элементы, на втором уровне – устойчивые комбинации элементов первого уровня и так далее вверх по цепочке. Причем сети сами формируют важные, по их мнению, элементы. Более детально алгоритм можно представить следующим образом [8]:

- загрузка учебного набора данных  $X_0$ ;
- определение параметров сети (количество слоев, размеры слоев, установление номера текущего слоя  $i=0$ );
- построение сети для слоев  $i, i+1$ ;
- обучение сети на наборе  $X_i$ ;
- удаление вспомогательного слоя сети;
- сохранение весов  $W_i$  связей слоев  $i, i+1$ ;
- генерация набора данных  $X_{i+1}$  для следующей сети (для этого необходимо пропустить через пару слоев  $i, i+1$  набор данных  $X_i$ );
- переход п. 3;
- конец работы.

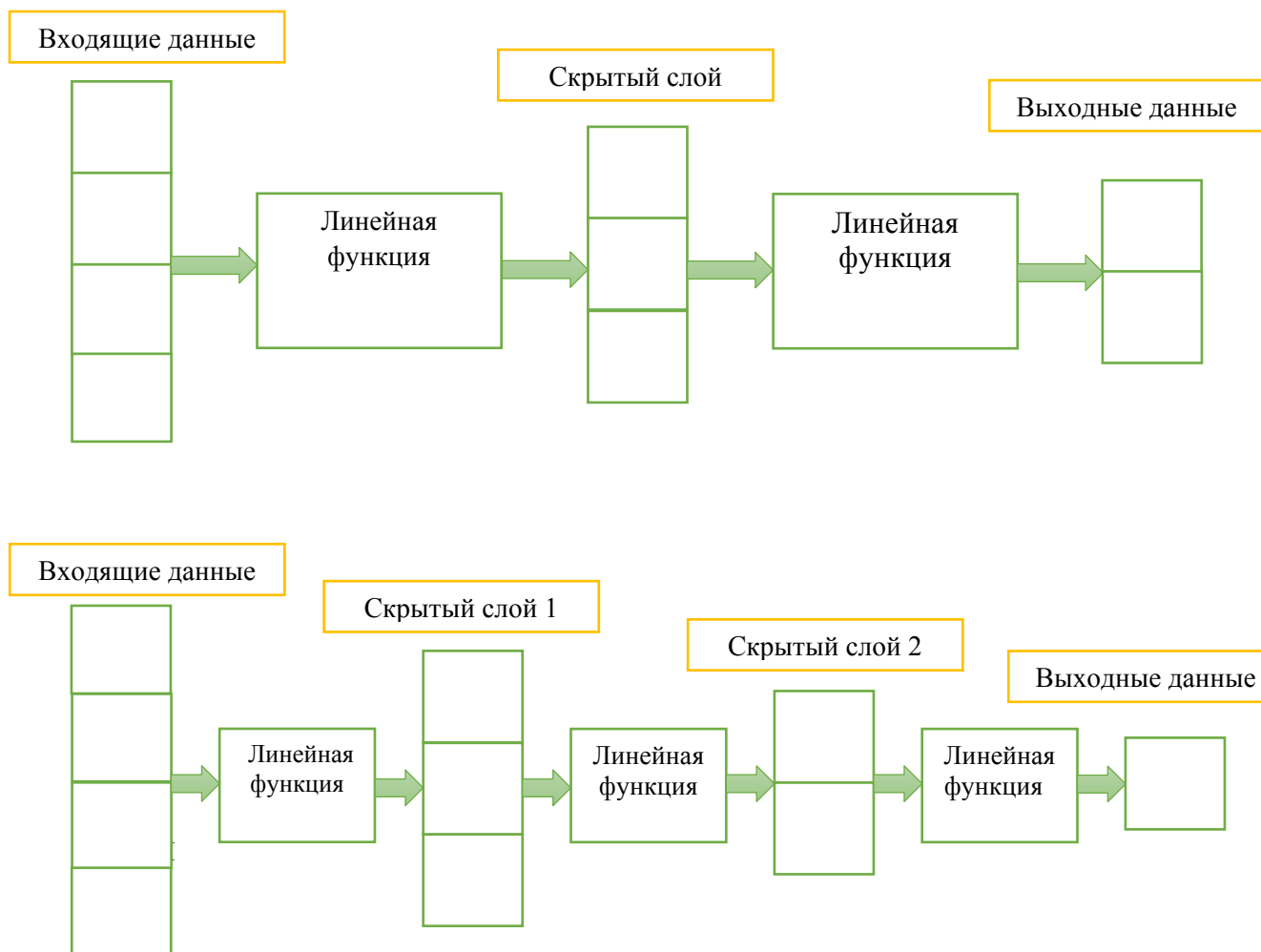


Рис. 1. Модификация способов обучения искусственной нейронной сети

Поэтому целью данной работы является модернизация уже имеющегося программного продукта «Нейропакет КДС 1.0» посредством внедрения в систему «глубокого обучения» для уменьшения погрешностей при прогнозировании пожароопасных показателей кислородсодержащих соединений.

В качестве объектов исследования выбраны органические соединения, широко применяемые в нефтегазовой отрасли. Так, метанол используется как ингибитор гидратообразования на газоконденсатных месторождениях; спиртобензол – для экстракции веществ кислотного характера битуминозных компонентов, силикагелевых смол и т.п.; спирты и простые эфиры – как добавки к бензинам; уксусная и муравьиная кислоты, сложные эфиры – как замедлители скорости растворения карбонатной породы; кетоны – для депарафинизации нефтепродуктов; альдегиды – в качестве биоцидов.

В качестве прогнозируемого пожароопасного свойства вещества выступает температура вспышки. Это один из распространенных показателей для исследования, поскольку обладает широким спектром применения в деятельности проектных организаций и надзорных органов.

Для описания строения молекул обучающей выборки исследуемых соединений применяются молекулярные дескрипторы структурной формулы: Wiener index, Randic index (order 0), Randic index (order 1), Randic index (order 2), Randic index (order 3), Kier&Hall index (order 0), Kier&Hall index (order 1), Kier&Hall index (order 2), Kier&Hall index (order 3), Kier shape index (order 1), Kier shape index (order 2), Kier shape index (order 3), Kier flexibility index, Average Information content (order 0), Information content (order 0), Average Structural

Information content (order 0), Structural Information content (order 0), Average Complementary Information content (order 0), Complementary Information content (order 0), Average Bonding Information content (order 0), Bonding Information content (order 0), Average Information content (order 1), Information content (order 1), Average Structural Information content (order 1), Structural Information content (order 1), Average Complementary Information content (order 1), Complementary Information content (order 1), Average Bonding Information content (order 1), Bonding Information content (order 1), Average Information content (order 2), Information content (order 2), Average Structural Information content (order 2), Structural Information content (order 2), Average Complementary Information content (order 2), Complementary Information content (order 2), Average Bonding Information content (order 2), Bonding Information content (order 2), Valaban index) и другие, которые наиболее точно описывают разницу в строении между структурными изомерами [9–11].

В ходе прогнозирования температуры вспышки исследуемых веществ была смоделирована искусственная нейронная сеть, представленная на рис. 2, а в таблице представлены полученные результаты.

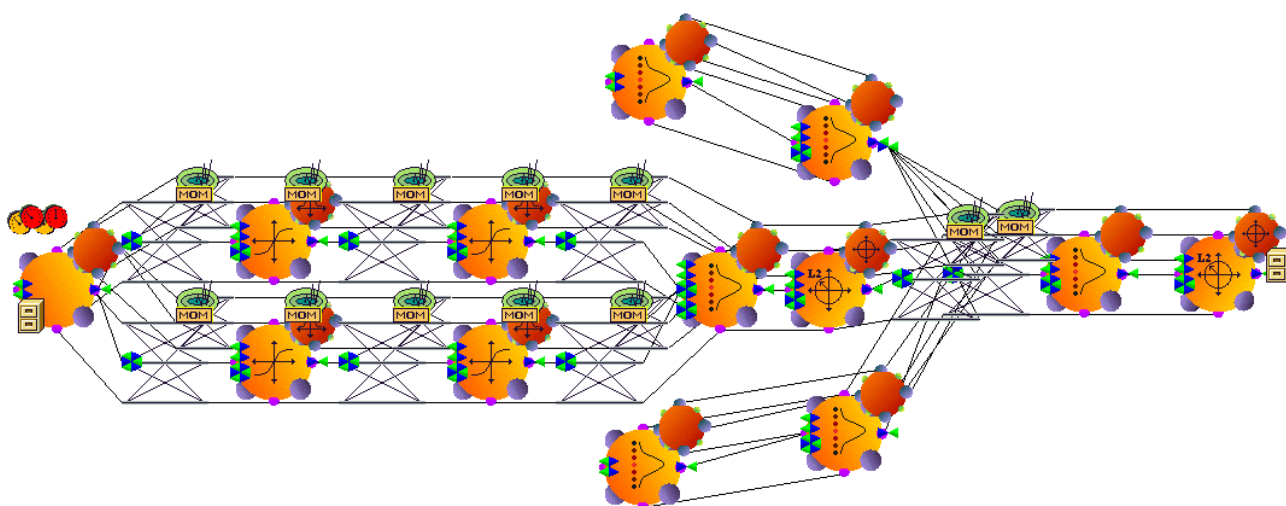


Рис. 2. Искусственная нейронная сеть с углубленным обучением

Анализируя полученные значения, видно, что при прогнозировании температуры вспышки предельных кетонов средняя относительная погрешность составила 3,35 %, а абсолютная погрешность – 1 °С, предельных альдегидов относительная и абсолютная погрешность составили 1,5 % и 1,5 °С соответственно, сложных эфиров масляной кислоты относительная погрешность не превысила 1,7 %, абсолютная погрешность – 1,5 °С.

Таким образом, оригинальная компьютерная программа «Нейропакет КДС 1.0» и внедренная в нее библиотека (Deep Learn Toolbox), обеспечивающая «глубокое обучение» искусственных нейронных сетей, позволяет более точно предсказывать показатель пожарной опасности, в частности температуру вспышки.

Таблица. Результаты прогнозирования усовершенствованным программным продуктом «Нейропакет КДС 1.0»

Вещество	Температура вспышки, °С		Относительная погрешность расчетов, %	Абсолютная погрешность, °С
	прогнозируемая	справочная [12, 13]		
Пределные кетоны				
3-Пентанон	13	–	–	–
3-Гексанон	32	–	–	–
2-Гептанон	46	–	–	–
3-Октанон	53	–	–	–
2-Октанон	59	–	–	–
3-Нонанон	70	–	–	–
2-Деканон	81	–	–	–
6-Ундеканон	93	–	–	–
2-Додеканон	107	–	–	–
Ацетон	-23	-22	4,5	1
Дипропилкетон	45	44	2,2	1
Бутилметилкетон	11	11	–	–
Средняя относительная и абсолютная погрешность			3,35	1
Пределные альдегиды				
2-Фенилпропаналь	70	70	–	–
2-Метилбензальдегид	80	79	1,2	1
2,5-Диметилбензальдегид	90	89	1,1	1
4-Этилбензальдегид	85	83	2,4	2
2,6-Диметилбензальдегид	98	97	1,0	1
3,5-Диметилбензальдегид	95	94	1,0	1
2,3-Диметилбензальдегид	102	101,7	0,2	0,3
4-Бутилбензальдегид	99	100	1,0	1
2,4,6-Триметилбензальдегид	101	102	0,9	1
2,4,5-Триметилбензальдегид	105	111	5,4	6
Средняя относительная и абсолютная погрешность			1,5	1,5
Сложные эфиры масляной кислоты				
Гексилбутират	180	179	0,5	1
Бутилбутират	154	–	–	–
Метилбутират	139	–	–	–
Амилбутират	120	–	–	–
Гептилбутират	99	100	1	1
Децилбутират	129	127	1,5	2
Изобутилбутират	53	51	3,9	2
Пропилбутират	79	–	–	–
Изопропилбутират	62	–	–	–
Изоамилбутират	55	–	–	–
Этилбутират	118	–	–	–
Средняя относительная и абсолютная погрешность			1,7	1,5

### Литература

1. Технический регламент о требованиях пожарной безопасности: Федер. закон от 22 июля 2008 г. № 123-ФЗ (одоб. Советом Федерации 11 июля 2008 г.) // Рос. газ. 2008. № 163.
2. Королев Д.С., Калач А.В., Каргашилов Д.В. Прогнозирование пожароопасных свойств веществ и материалов с использованием дескрипторов и нейронных сетей // Науч.-теоретич. журн. «Вестник БГТУ им. В.Г. Шухова». 2015. № 4. С. 100–103.
3. Королев Д.С., Калач А.В., Рудаков О.Б. Прогнозирование пожароопасных свойств веществ // Безопасность в техносфере. 2015. Т. 56. № 5. С. 3–6.

4. Королев Д.С., Калач А.В., Каргашилов Д.В. Свидетельство о государственной регистрации программы для ЭВМ № 2016614070 «Нейропакет КДС 1.0»; правообладатель ФГБОУ ВО Воронежский ин-т ГПС МЧС России. № 2016611455; заяв. 24.02.16; зарегистрировано в реестре программ для ЭВМ 16.04.16.

5. Cybernetic methods of drug design. I. Statement of the problem the perceptron approach / S.A. Hiller [et al.] // *Comput. Biomed. Res.* 1973. V. 6. № 5. P. 411–421.

6. Zupan J., Gasteiger J. Neural networks: a new method for solving chemical problems or just a passing phase? // *Anal. Chim. Acta.* 1991. V. 248. № 1. С. 1–30.

7. Freeman J.A., Skapura D.M. Neural networks: algorithms, applications and programming techniques. Addison-Wesley Publishing Company: Menlo Park, CA, 1991. 414 p.

8. Сайт Deep learning (глубокое или глубинное обучение). URL: <http://www.insycom.ru/html/metodmat/dp.pdf> (дата обращения: 20.11.2016).

9. Ngoc L.M., Yoon- Mo K. Quantitative prediction of lipase reaction in ionic liquids by QSAR using COSMO-RS molecular descriptors // *Biochemical Engineering Journal.* 2014. Vol. 87. № 5. Pp. 33–40.

10. Free S.M., Jr., Wilson J.W. A Mathematical Contribution to Structure-Activity Studies // *J. Med. Chem.* 1964. V. 7. № 4. P. 395–399.

11. Golender V.E., Rozenblit A.B. Logico-structural approach to computer-assisted drug design // *Med. Chem. (Academic Press).* 1980. V. 11. № 9. P. 299–337.

12. Корольченко А.Я., Корольченко Д.А. Пожаровзрывоопасность веществ и материалов и средства их тушения: справ. в 2-х ч. 2-е изд., перераб. и доп. М.: Асс. «Пожнаука», 2004. Ч. I. 713 с.

13. Корольченко А.Я., Корольченко Д.А. Пожаровзрывоопасность веществ и материалов и средства их тушения: справ. в 2-х ч. 2-е изд., перераб. и доп. М.: Асс. «Пожнаука», 2004. Ч. II. 774 с.

## References

1. Technical regulations for fire safety requirements. Federal Law on 22.07.2008 No. 123. Ros. gaz. 2008. № 163. (in Russian).

2. Korolev D.S., Kalach A.V., Kargashilov D.V. Prognozirovanie požaroopasnih svoystv veschestv i materialov s ispolzovaniem deskriptorov i neyronnyh setey [Prediction of properties of substances and flammable materials using descriptors and neural networks] *Vestnik BGTU im. Shukhova. Bulletin of BSTU named after V.G. Shukhov*, 2015. no. 4. Pp. 100–103.

3. Korolev D.S., Kalach A.V., Rudakov O.B. Prognozirovaniye požaroopasnykh svoystv veshchestv [Forecasting of fire-dangerous properties of substances]. *Bezopasnost v tehnosfere – Safety in a technosphere.* 2015. vol. 56. № 5. Pp. 3–6.

4. Korolev D.S., Kalach A.V., Kargashilov D.V. Svidetelystvo o gosudarstvennoy registraciy programmy dlay EVM № 2016614070 «Neuropackage KDS 1.0». rightholder: the Voronezh Institute of the State Fire Service. № 2016611455; stated 24.02.16; registered in the registry of the computer programs 16.04.16.

5. Cybernetic methods of drug design. I. Statement of the problem the perceptron approach / S.A. Hiller [et al.] // *Comput. Biomed. Res.* 1973. V. 6. № 5. P. 411–421.

6. Zupan J., Gasteiger J. Neural networks: a new method for solving chemical problems or just a passing phase? // *Anal. Chim. Acta.* 1991. V. 248. № 1. С. 1–30.

7. Freeman J.A., Skapura D.M. Neural networks: algorithms, applications, and programming techniques. – Addison-Wesley Publishing Company: Menlo Park, CA. 1991. 414 p.

8. Site Deep learning (Deep or deep learning) [electronic resource] access mode. URL: <http://www.insycom.ru/html/metodmat/dp.pdf> (date of the application 20.11.2016).

9. Ngoc L.M., Yoon- MoK. Quantitative prediction of lipase reaction in ionic liquids by QSAR using COSMO-RS molecular descriptors // *Biochemical Engineering Journal.* 2014.

Vol. 87. № 5. Pp. 33–40.

10. Free S.M., Jr., Wilson J.W. A Mathematical Contribution to Structure-Activity Studies. // J. Med. Chem. 1964. V. 7. № 4. P. 395–399.

11. Golender V.E., Rozenblit A.B. Logico-structural approach to computer-assisted drug design // Med. Chem. (Academic Press). 1980. V. 11. № 9. P. 299–337.

12. Korolchenko A.Ya., Korolchenko D.A. Pozharovzryvoopasnost veshchestv i materialov i sredstva ikh tusheniya: sprav. 2-e izd. [Fire and explosion hazard of substances and materials and their means of fighting. Reference. 2nd ed.]. Moscow, Pozhnauka Publ., 2004. Part I. 713 p.

13. Korolchenko A.Ya., Korolchenko D.A. Pozharovzryvoopasnost veshchestv i materialov i sredstva ikh tusheniya: sprav. 2-e izd. [Fire and explosion hazard of substances and materials and their means of fighting. Reference. 2nd ed.]. Moscow, Pozhnauka Publ., 2004. Part II. 774 p.