

ИСКУССТВЕННАЯ НЕЙРОННАЯ СЕТЬ И МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ И КЛАССИФИКАЦИИ УСТОЙЧИВОСТИ КОРОНАТОВ НАТРИЯ И КАЛИЯ

© 2019 г. Н. В. Бондарев*

*Харьковский национальный университет имени В. Н. Каразина,
пл. Свободы 4, Харьков, 61022 Украина
e-mail: bondarev_n@rambler.ru

Поступило в Редакцию 23 января 2019 г.

После доработки 23 января 2019 г.

Принято к печати 10 февраля 2019 г.

Разработаны модели множественной линейной регрессии и многослойной искусственной нейронной сети для моделирования и прогнозирования констант устойчивости коронатов натрия и калия по свойствам водно-органических растворителей вода–метанол, вода–пропан-2-ол, вода–ацетонитрил, вода–ацетон. Приведены спрогнозированные значения констант устойчивости коронатов в водно-этанольных растворителях и проведено сравнение результатов моделей множественной линейной регрессии и искусственной нейронной сети. На основе моделей множественной линейной регрессии и принципа линейности свободных энергий проведена количественная оценка вкладов электростатических, когезионных и электронодонорных взаимодействий в возрастание устойчивости коронатов. Построены модели нейронной сети для классификации устойчивости коронатов натрия и калия на основе алгоритмов обучения без учителя (многослойные персептроны) и с учителем (сети Кохонена). Показано, что нейросетевые классификаторы полностью подтвердили классификацию устойчивости коронатов, проведенную разведочным методом *k*-средних.

Ключевые слова: множественная регрессия, многослойный персептрон, сеть Кохонена, прогнозирование, нейросетевой классификатор

DOI: 10.1134/S0044460X1907014X

Классические методы прогнозирования в большинстве своем основываются на предположении о линейной зависимости между прогнозируемой (зависимой) переменной и множеством независимых дескрипторов. Для построения моделей линейной множественной регрессии используются методы корреляционного и регрессионного анализа.

Нелинейные зависимости между переменными традиционно моделируются методами подгонки кривых (квадратичных, кубических, степенных, показательных, логарифмических, гиперболических, экспоненциальных, логистических), методами преобразования нелинейных соотношений в линейную форму и другими. Однако в последние десятилетия более перспективным методом моделирования нелинейных зависимостей в задачах прогнозирования является использование искусственных

нейронных сетей [1]. Основные преимущества алгоритмов искусственной нейронной сети проявляются в способности сетей к обучению, обобщению и прогнозированию данных, отказоустойчивости и параллельной обработке информации в дополнение к быстрой вычислительной процедуры. Подтверждением этому являются научные разработки с приложениями нейронных сетей в теоретической и вычислительной химии, аналитической химии, биохимии, медицине, лекарственной химии, в фармации и исследованиях пищевых продуктов.

Следует отметить, что искусственные нейронные сети стали применять к решению хеометрических проблем в начале 90-х годов XX столетия. Подробно описаны четыре приложения для химической инженерии [2]: обнаружение

неисправностей, прогнозирование качества, обработка сигналов, моделирование и контроль процессов. Представлены различные архитектуры искусственных нейронных сетей и их применение в химии [3], отмечены плюсы и минусы искусственной нейронной сети по сравнению с традиционными хемометрическими методами. Предложен новый подход [4], так называемая искусственная нейронная сеть с физикой и химией (Phys-Chem ANN), для предсказания биоактивности пептидов и белков. В основу подхода ANN Phys-Chem положены физические и химические свойства, а также структурные особенности белков. Решена задача классификации и прогнозирования силы слабых органических кислот в водно-органических растворителях [5].

С помощью нейросетевых технологий впервые проведено моделирование одиночных волновых кинетических кривых [6], показана важность химических сдвигов ^1H , ^{13}C и ^{15}N белков для подтверждения и уточнения трехмерной структуры белка [7], разработан алгоритм спектроскопических исследований следовых количеств газов [8], предсказана липофильность химических веществ [9, 10], реализован метод прогнозирования времени дрейфа ионов пептидов в масс-спектрометрии ионов [11], разработаны модели прогнозирования теплотворной способности словенского угля [12], предложен метод одновременного анализа двух компонентов порошкообразного лекарственного препарата парацетамола и дифенгидрамина гидрохлорида по данным ИК спектроскопии [13], выполнен количественный анализ порошка калия феноксиметилпенициллина по данным ближней инфракрасной спектроскопии [14], проанализировано влияние демографических, биохимических параметров и терапевтических характеристик пациентов на концентрацию сыворотки топирамата [15], построены модели прогнозирования трипановидной активности хиноноидных соединений [16], исследована противоопухолевая (противораковая) активность фенольных соединений [17], разрабатываются новые лекарства [18], исследована психоактивность каннабиноидных соединений [19], выявлена взаимосвязь между фенольными соединениями и антиоксидантной активностью пищевых продуктов [20]. На основе нейронных сетей разработаны: новая технология для интеллектуальных датчиков газа, работающих в динамичной среде [21], алгоритмы классификации качества картофеля [22], вин [23–25], меда [26, 27],

сыра [28], чая [29], оливковых [30] и растительных масел [31].

Искусственные нейронные сети успешно применяются для интерпретации спектров [32, 33], оптимизации медицинской диагностики патологических заболеваний [34], моделирования свойств рыбных антибиотиков [35]. Также искусственные нейронные сети используются для прогнозирования константы скорости реакции взаимодействия химических соединений с гидроксильной группой [36], скорости растворения колеманита в воде, насыщенного углекислым газом [37], плазмонов в наностержнях серебра [38], потенциалов ионизации атомов I–III групп периодической системы Д.И. Менделеева [39], выживаемости пациентов с травмой [42], активности каннабиноидных лигандов [40] и органических реакций [41] по отпечаткам пальцев.

Данная статья посвящена моделированию, прогнозированию и классификации устойчивости коронатов натрия и калия в водно-органических растворителях методами нейронных сетей и множественной линейной регрессии. Константы устойчивости коронатов натрия и калия в водно-органических растворителях (вода–метанол, вода–пропан-2-ол, вода–ацетонитрил, вода–ацетон) получены ранее и представлены в монографии [43]. Там же приведены литературные данные по свойствам водно-органических растворителей: диэлектрическая проницаемость, параметры Димрота–Райхардта E_T и Камлета–Тафта B_{KT} , плотность энергии когезии δ^2 .

Построение и анализ результатов моделей множественной линейной регрессии и нейросетевых моделей проведены с использованием статистического пакета STATISTICA 12 в среде Windows 10.

Множественная линейная регрессия. В табл. 1 приведены описательные статистики для исходных данных [матрица данных размером $n \times m$, $n = 33$ – количество наблюдений; $m = 6$ – количество зависимых ($\lg K_{18C6Na^+}$, $\lg K_{18C6K^+}$) и независимых параметров (четыре свойства водно-органических растворителей)]. Для проверки характера распределения переменных на нормальность рассчитан критерий Колмогорова–Смирнова (табл. 2). Расчетные значения критерия для всех переменных меньше табличного $D_{\text{табл}} = 0.2308$ для $n = 33$ и $\alpha = 0.05$. Это говорит о том, что эмпирические распределения переменных не отличаются от

Таблица 1. Описательные статистики исходных данных ($n = 33$)

Параметр	Среднее значение	Минимальное значение	Максимальное значение	Дисперсия	Стандартное отклонение	Стандартная ошибка среднего
$\lg K_{18C6Na^+}$	2.17	0.52	4.25	1.15	1.07	0.19
$\lg K_{18C6K^+}$	3.50	2.04	5.85	1.02	1.01	0.18
$1/\varepsilon$	2.00	1.00	4.12	0.82	0.91	0.16
E_T	0.75	0.46	1.00	0.02	0.13	0.02
$B_{КТ}$	0.51	0.19	0.89	0.04	0.19	0.03
δ^2	0.67	0.24	1.00	0.05	0.23	0.04

Таблица 2. Расчетные значения критерия Колмогорова–Смирнова

Параметр	Расчетное значение критерия	Параметр	Расчетное значение критерия
$\lg K_{18C6Na^+}$	0.0987	E_T	0.1230
$\lg K_{18C6K^+}$	0.1383	$B_{КТ}$	0.1250
$1/\varepsilon$	0.1773	δ^2	0.0782

Таблица 3. Результаты корреляционного анализа

Параметр	Среднее	Стандартное отклонение	$1/\varepsilon$	E_T	$B_{КТ}$	δ^2	$\lg K_{18C6Na^+}$	$\lg K_{18C6K^+}$
$1/\varepsilon$	2.00	0.91	1.000	-0.827	0.899	-0.641	0.352	0.511
E_T	0.75	0.13	-0.827	1.000	-0.763	0.732	-0.494	-0.648
$B_{КТ}$	0.51	0.19	0.899	-0.763	1.000	-0.507	0.340	0.406
δ^2	0.67	0.23	-0.641	0.732	-0.507	1.000	-0.849	-0.970
$\lg K_{18C6Na^+}$	3.50	1.01	0.352	-0.494	0.340	-0.849	1.000	0.928
$\lg K_{18C6K^+}$	2.17	1.07	0.511	-0.648	0.406	-0.970	0.928	1.000

нормального. Следовательно, множественно-регрессионный анализ исследуемых процессов комплексообразования катионов натрия и калия с 18-краун-6 эфиром представляется правомерным.

В табл. 3 приведены результаты корреляционного анализа переменных. Константы устойчивости коронатов ($\lg K_{18C6Na^+}$, $\lg K_{18C6K^+}$) демонстрируют сильную отрицательную корреляцию с плотностью энергии когезии, средняя корреляция зависимой переменной $\lg K_{18C6Na^+}$ наблюдается с переменными $1/\varepsilon$ и E_T , самая слабая корреляция с дескрипторами $1/\varepsilon$ и $B_{КТ}$ у переменной $\lg K_{18C6K^+}$.

При отборе переменных в регрессионную модель использованы два метода, представленных

в пакете STATISTICA 12. Это метод прямого выбора регрессоров и метод обратного их исключения. В табл. 4 показаны результаты множественной линейной регрессии. Значения коэффициента детерминации $R^2 = 0.961$ для модели $\lg K_{18C6Na^+}$ и $R^2 = 0.835$ для модели $\lg K_{18C6K^+}$ показывают, что линейная регрессия объясняет 96.1% и 83.5% разброса значений переменных $\lg K_{18C6Na^+}$ и $\lg K_{18C6K^+}$ относительно среднего уровня. Это свидетельствует о значительной степени зависимости констант устойчивости коронатов натрия и калия от выбранных для анализа свойств (регрессоров) водно-органических растворителей.

Значения F -критерия Фишера $F_{расч}(2, 30) = 374.23$ (при степенях свободы $\nu_1 = 2$, $\nu_2 = 30$) и

Таблица 4. Результаты множественной регрессии

Параметр	B_i	Стандартная ошибка B_i	$ t _{\text{расч}}(30)$	p -Уровень
Модель $\lg K_{18C6Na^+} = B_1 + B_2\delta^2 + B_31/\varepsilon$ $R = 0.981, R^2 = 0.961, F_{\text{расч}}(2, 30) = 374.23, p = 0.00$				
B_1	6.08	0.24	25.53	0.00
δ^2	-5.18	0.22	23.35	0.00
$1/\varepsilon$	-0.22	0.06	4.03	0.00
Параметр	B_i	Стандартная ошибка B_i	$ t _{\text{расч}}(29)$	p -Уровень
Модель $\lg K_{18C6K^+} = B_1 + B_2\delta^2 + B_31/\varepsilon + B_4B_{KT}$ $R = 0.914, R^2 = 0.835, F_{\text{расч}}(3, 29) = 48.75, p = 0.00$				
B_1	7.31	0.47	15.45	0.00
δ^2	-5.01	0.45	11.14	0.00
$1/\varepsilon$	-0.93	0.22	4.23	0.00
B_{KT}	2.76	0.93	2.97	0.01

$F_{\text{расч}}(3, 29) = 48.75$ (при степенях свободы $\nu_1 = 3$, $\nu_2 = 29$) говорят об адекватности регрессионных моделей. Критические значения при уровне значимости $p = 0.05$ для данных степеней свободы $F_{\text{кр}}(2, 30) = 3.32$, $F_{\text{кр}}(3, 29) = 2.93$; $F_{\text{расч}} > F_{\text{кр}}$.

Расчетные значения критерия Стьюдента для коэффициентов множественной линейной регрессии $|t|_{\text{расч}}(30)$ в пределах от 4.03 до 25.53 при степенях свободы $\nu = 30$ и $|t|_{\text{расч}}(29)$ в пределах от 2.97 до 15.45 при степенях свободы $\nu = 29$ позволяют заключить, что коэффициенты B_i регрессионных моделей (модель $\lg K_{18C6Na^+}$, модель $\lg K_{18C6K^+}$) являются значимыми с доверительной вероятностью 95% по критерию Стьюдента. Критические значения при уровне значимости $p = 0.05$ для соответствующих степеней свободы $t_{\text{кр}}(30) = 2.04$, $t_{\text{кр}}(29) = 2.05$; $t_{\text{расч}} > t_{\text{кр}}$.

Из данных табл. 4 получаем следующие регрессионные модели:

$$\lg K_{18C6Na^+} = 6.08 \pm 0.49 - (5.18 \pm 0.45)\delta^2 - (0.22 \pm 0.12)1/\varepsilon, \quad (1)$$

$$\lg K_{18C6K^+} = 7.31 \pm 0.96 - (5.01 \pm 0.92)\delta^2 - (0.93 \pm 0.45)1/\varepsilon + (2.76 \pm 1.91)B_{KT}. \quad (2)$$

Для доказательства правильности построенных регрессионных моделей проведен анализ остатков по критерию Колмогорова–Смирнова. Значения расчетного критерия для модели $\lg K_{18C6Na^+}$ $D_{\text{расч}} = 0.156$, для модели $\lg K_{18C6K^+}$ $D_{\text{расч}} = 0.154$. Расчетные значения критерия $D_{\text{расч}}$ меньше критического $D_{\text{кр}} =$

0.231 для $n = 33$ и $p = 0.05$. Следовательно, есть основания полагать, что остатки регрессионных моделей подчиняются закону нормального распределения, т. е. модели адекватно описывают зависимость констант устойчивости коронатов от свойств растворителя.

Сопоставление результатов сольватационно-термодинамического анализа, проведенного нами ранее [43] (рис. 1), и множественного регрессионного анализа (рис. 2) позволяет раскрыть природу взаимодействий в системах водно-органический растворитель–соль (NaCl или KCl)–краун-эфир, энергетика которых обуславливает рост устойчивости коронатов натрия и калия с увеличением содержания органического компонента в смешанном растворителе.

Согласно сольватационно-термодинамической модели, возрастание устойчивости коронатов натрия и калия ($\Delta_{\text{tr}}G_{\text{LM}}^{\circ} < 0$) обусловлено понижением стабилизации сольватов катионов ($\Delta_{\text{tr}}G_{\text{M}}^{\circ} > 0$), краун-эфира ($\Delta_{\text{tr}}G_{\text{L}}^{\circ} > 0$) и ростом стабилизации образующихся сольватов коронатов ($\Delta_{\text{tr}}G_{\text{LM}}^{\circ} < 0$) при замене воды на водно-метанольные растворители и чистый метанол (рис 1).

Из анализа результатов моделей множественной регрессии можно заключить (рис. 2), что возрастание устойчивости коронатов натрия в водно-метанольных растворителях обусловлено энергетикой сил когезии [$\Delta\Delta G_{\text{LM}}^{\circ}(\delta^2) < 0$], в то время как проявление аналогичного эффекта для

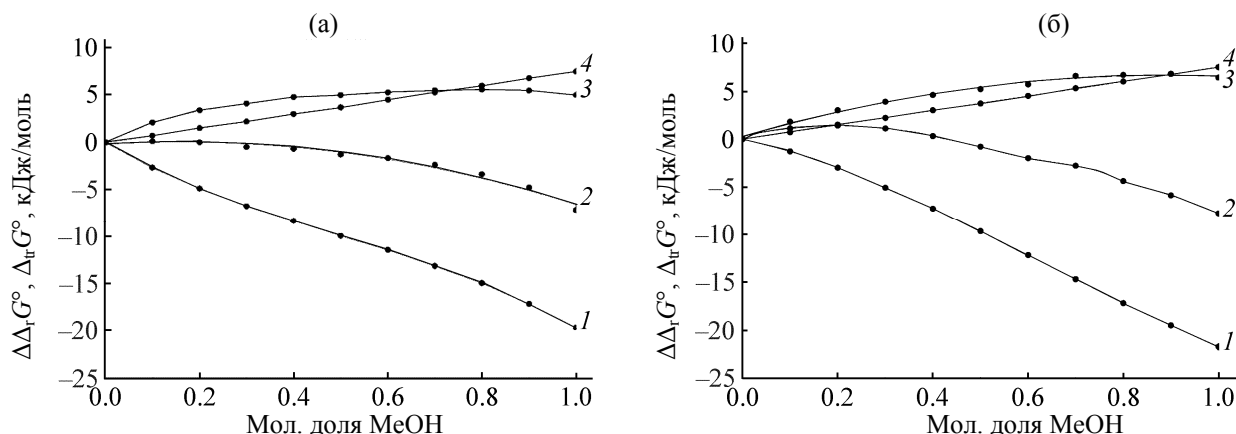


Рис. 1. Энергии Гиббса переноса реакции комплексообразования (1 – $\Delta\Delta_r G_{LM}^\circ$), реагентов (3 – $\Delta_r G_L^\circ$, 4 – $\Delta_r G_M^\circ$) и продуктов (2 – $\Delta_r G_{LM}^\circ$) реакции из воды в водно-метанольные растворители и метанол. (а) – 18-краун-6 Na^+ , (б) – 18-краун-6 K^+ . $\Delta\Delta_r G_{LM}^\circ = \Delta_r G_{LM}^\circ - \Delta_r G_L^\circ - \Delta_r G_M^\circ$ [43], где $M = \text{Na}^+$ или K^+ .

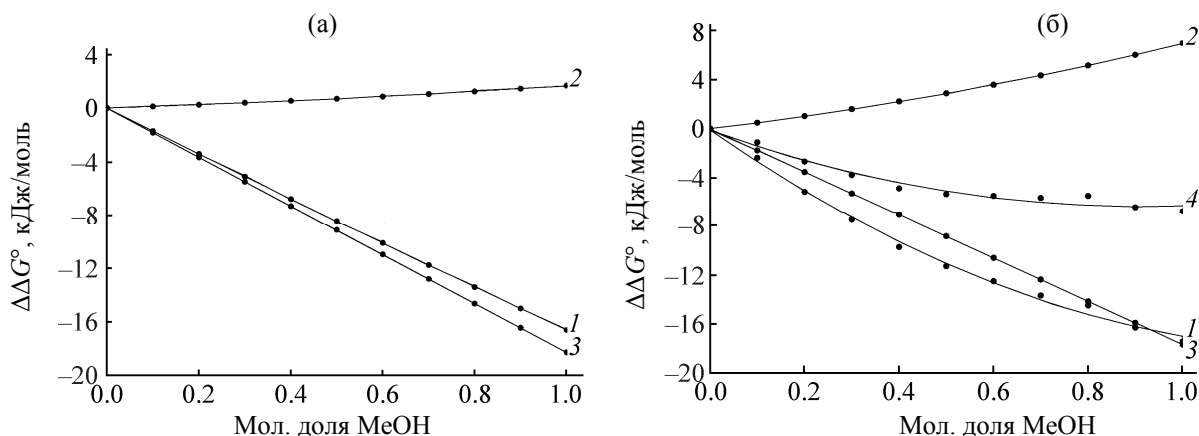


Рис. 2. Энергия Гиббса переноса реакции комплексообразования (1 – $\Delta\Delta_r G_{LM}^\circ$) и ее составляющих [2 – $\Delta\Delta G_{LM}^\circ(1/\epsilon)$, 3 – $\Delta\Delta G_{LM}^\circ(\delta^2)$, 4 – $\Delta\Delta G_{LM}^\circ(B_{KT})$] от состава водно-метанольного растворителя. (а) – 18-краун-6 Na^+ , (б) – 18-краун-6 K^+ . $\Delta\Delta_r G_{LM}^\circ = \Delta\Delta G_{LM}^\circ(1/\epsilon) + \Delta\Delta G_{LM}^\circ(\delta^2) + \Delta\Delta G_{LM}^\circ(B_{KT})$, где $M = \text{Na}^+$ или K^+ .

короната калия связано как с энергетикой когезионных взаимодействий [$\Delta\Delta G_{LM}^\circ(\delta^2) < 0$], так и с энергетикой электронодонорных взаимодействий [$\Delta\Delta G_{LM}^\circ(B_{KT}) < 0$]. Электростатические эффекты среды с ростом содержания метанола в смеси понижают устойчивость коронатов натрия и калия по сравнению с водой [$\Delta\Delta G_{LM}^\circ(1/\epsilon) > 0$].

Выявленные эффекты можно объяснить следующим образом. Уменьшение плотности энергии когезии в водно-метанольных растворителях по сравнению с водой облегчает образование полостей в растворителе, которые в большей мере стабилизируют сольваты коронатов натрия и калия. С уменьшением диэлектрической проницаемости растворителя возрастает межмолекулярное электростатическое взаимодействие ($\text{Na}^+ - \text{Cl}^-$ и $\text{K}^+ - \text{Cl}^-$) и, как следствие, снижается комплексо-

образующая способность пары $M-L$ (конкуренция за катион). Замена воды на водно-метанольные смеси повышает электронодонорную способность среды, что проявляется в дополнительной стабилизации сольватов короната калия водно-метанольным растворителем.

Нейросетевой и множественной линейной регрессии аппроксиматоры. Из построенных в Statistica Neural Networks [44, 45] тысячи нейронных сетей (сеть радиального базиса, многослойный персептрон) выбрана оптимальная архитектура сети по статистическим показателям – персептрон МП 4-7-2. Сеть имеет три слоя: входной слой с четырьмя нейронами, выходной – два нейрона и один скрытый слой с семью нейронами (табл. 5).

Таблица 5. Статистические показатели персептрона МП 4-7-2^a

Подвыборка	МП 4-7-2		Коэффициент корреляции	
	производительность	ошибка обучения	$\lg K$ 18-краун-6Na ⁺	$\lg K$ 18-краун-6K ⁺
Обучающая	0.9978	0.0038	0.9971	0.9985
Контрольная	0.9997	0.0099	0.9995	0.9999
Тестовая	0.9993	0.0077	0.9992	0.9995

^a Производительность обучения, контрольная производительность, тестовая производительность – отношение стандартного отклонения ошибки прогноза к стандартному отклонению исходных данных на соответствующих выборках. Ошибка обучения, контрольная ошибка, тестовая ошибка – ошибка сети на соответствующих выборках.

Таблица 6. Результаты аппроксимации зависимости констант устойчивости коронатов натрия и калия от свойств смешанного растворителя вода–ацетонитрил

Мол. доля CH ₃ CN	$\lg K(18\text{-краун-6Na}^+)$			$\lg K(18\text{-краун-6K}^+)$		
	$\lg K_{\text{эксп}}$	множественная линейная регрессия	многослойный персептрон	$\lg K_{\text{эксп}}$	множественная линейная регрессия	многослойный персептрон
Вода	0.52	0.68	0.63	2.04	1.89	2.14
0.1	1.18	1.03	1.06	2.42	2.57	2.41
0.2	1.46	1.39	1.33	2.74	2.95	2.67
0.3	1.55	1.74	1.61	3.02	3.24	2.96
0.4	1.64	2.09	1.86	3.28	3.46	3.23
0.5	2.17	2.45	2.19	3.53	3.71	3.51
0.6	2.69	2.81	2.66	3.80	4.04	3.83
0.7	3.20	3.17	3.17	4.08	4.36	4.15
0.8	3.65	3.53	3.64	4.41	4.60	4.36
0.9	4.01	3.90	4.00	4.78	4.81	4.80
1.0	4.25	4.26	4.26	5.20	5.01	5.18
Средняя ошибка аппроксимации, %		10.2	5.5		5.6	1.4

Трехслойный персептрон обучен [46] на основе одношагового квазиньютоновского алгоритма BFGS (Бройдена–Флетчера–Гольдфарба–Шанно) [47]. Обучение завершено на 85 эпохе с ошибкой обучения 0.0038, контрольной ошибкой 0.0099 и тестовой ошибкой 0.0077. Ошибки получены соответственно для обучающего (70%), контрольного (15%) и тестового (15%) подмножеств, на которые был разбит исходный массив констант устойчивости коронатов и свойств водно-органических растворителей. Данные в обучающей выборке использовались для обучения сети.

Наблюдения в контрольной выборке – для проведения проверки качества обучения сети, чтобы исключить переобучение и определить момент остановки обучения сети. Данные в тестовой выборке не использовались при обучении сети, но обученная сеть применялась к этим данным для независимой проверки качества сети. Остановка обучения сети осуществлялась по достижении минимальной ошибки на контрольном подмножестве [44–46]. Ошибка обучения нейронной сети представляет собой суммарное квадратичное отклонение, т.е. среднеквадратичное отклонение

Таблица 7. Прогностический потенциал множественной линейной регрессии и аппроксиматоров многослойных персептронов

Мол. доля ацетона	$\lg K(18\text{-краун-}6\text{Na}^+)$			$\lg K(18\text{-краун-}6\text{K}^+)$		
	$\lg K_{\text{эксп}}$	множественная линейная регрессия	многослойный персептрон	$\lg K_{\text{эксп}}$	множественная линейная регрессия	многослойный персептрон
Вода	0.52	0.68	0.64	2.04	1.90	2.14
0.1	1.29	1.05	1.01	2.53	2.66	2.44
0.2	1.77	1.41	1.28	2.99	3.08	2.81
0.3	2.04	1.77	1.51	3.41	3.27	3.18
0.4	2.24	2.13	1.76	3.80	3.38	3.44
0.5	2.50	2.48	2.17	4.13	3.52	3.54
0.55	2.68	2.65	2.49	4.29	3.65	3.54
Средняя ошибка прогноза, %		12.8	20.0		8.6	7.4

предсказаний сети от эмпирических входных данных по константам устойчивости коронатов. Нейроны скрытого и выходного слоев активировались, соответственно, логистической и тождественной активационными функциями.

Одним из способов проверки качества аппроксимации является сравнение наблюдаемых значений $\lg K_{\text{эксп}}$ и предсказанных результатов $\lg K_{\text{расч}}$ методами множественной линейной регрессии и обученным персептроном МП 4-7-2 (табл. 6). Из анализа данных табл. 6 можно заключить, что оба метода проявляют достаточно высокую аппроксимирующую способность, однако нейронная сеть более эффективна, она имеет меньшую ошибку регрессии.

Этот вывод также подтвержден результатами применения моделей множественной линейной регрессии и многослойных персептронов для аппроксимации констант устойчивости коронатов натрия и калия по свойствам водно-ацетоновых растворителей (табл. 7). Важно отметить, что эти данные не использовались при построении базовых прогностических моделей множественной регрессии и при обучении нейронных сетей, т. е. они являются основой для независимого тестирования множественной линейной регрессии и аппроксиматоров многослойных персептронов.

В табл. 8 приведены константы комплексообразования 18-краун-6 эфира с катионами Na^+ и

K^+ , полученные с помощью построенных аппроксиматоров множественной линейной регрессии и многослойных персептронов, на основе свойств водно-этанольных растворителей (диэлектрическая проницаемость, плотность энергии когезии, параметры Димрота–Райхардта и Камлета–Тафта). Следует заметить, что в литературе отсутствуют данные по константам устойчивости монокоронатов натрия и калия в смесях вода–этанол.

Нейросетевые классификаторы. Ранее была проведена кластеризация устойчивости коронатов натрия и калия методом k -средних [48]. Выполненный кластерный анализ данных позволил интерпретировать содержимое трех кластеров (классов).

Кластер 1 – умеренно устойчивые коронаты в смешанных растворителях промежуточного состава ($\lg K_{\text{NaL}} = 1.5\text{--}2.5$, 10 констант устойчивости; $\lg K_{\text{KL}} = 3.1\text{--}3.9$, 7 констант устойчивости).

Кластер 2 – устойчивые комплексы в растворителях с большим содержанием органического компонента и в чистых неводных растворителях ($\lg K_{\text{NaL}} = 2.6\text{--}4.3$, 12 констант устойчивости; $\lg K_{\text{KL}} = 4.0\text{--}5.2$, 10 констант устойчивости).

Кластер 3 – слабо устойчивые коронаты в воде и смешанных растворителях с большим содержанием воды ($\lg K_{\text{NaL}} = 0.5\text{--}1.4$, 11 констант устойчивости; $\lg K_{\text{KL}} = 2.0\text{--}3.0$, 16 констант устойчивости).

Таблица 8. Прогнозируемые значения констант устойчивости коронатов натрия и калия в водно-этанольных растворителях

Мол. доля EtOH	lgK(18-краун-6Na ⁺)		lgK(18-краун-6K ⁺)	
	множественная линейная регрессия	многослойный персептрон	множественная линейная регрессия	многослойный персептрон
Вода	0.52	0.52	2.04	2.04
0.1	1.00	0.86	2.40	2.32
0.2	1.32	1.38	3.11	2.80
0.3	1.63	1.66	3.41	3.24
0.4	1.94	1.92	3.61	3.74
0.5	2.25	2.22	3.82	4.27
0.6	2.56	2.61	4.03	4.70
0.7	2.88	3.06	4.23	4.82
0.8	3.19	3.56	4.47	4.73
0.9	3.51	4.01	4.77	4.76
1.0	3.83	4.27	4.95	4.72

Таблица 9. Основные характеристики классификаторов МП 4-4-3 и МП 4-3-3^a

Алгоритм обучения	Функция ошибки	Функция активации нейронов		Количественный состав кластеров		
		скрытый слой	выходной слой	класс 1	класс 2	класс 3
МП 4-4-3 18-краун-6Na ⁺						
BFGS 42	SOS	Logistic	Tanh	10	12	11
МП 4-4-3 18-краун-6K ⁺						
BFGS 11	Entropy	Logistic	Softmax	7	10	16

^a SOS – среднеквадратичная ошибка $E = 1/P[\sum_{k=1}^P (\lg K_{\text{расч},k} - \lg K_{\text{эксп},k})^2]$, P – число примеров в выборке; Entropy – кросс-энтропия [51] $H(p, q) = -\sum_x p(x) \log q(x)$, p и q – несвязанные друг с другом случайные переменные; Logistic – логистическая функция $\sigma(x) = 1/[1 + \exp(-x)]$, $[0, 1]$; Tanh – гиперболический тангенс $\text{th}(Ax) = (e^{Ax} - e^{-Ax})/(e^{Ax} + e^{-Ax})$, $[-1, 1]$; Softmax – обобщение логистической функции для многомерного случая $\sigma(z_j) = e^{z_j}/\sum_{k=1}^K e^{z_k}$, $[0, 1]$.

Данная классификация была подтверждена [48] дискриминантным анализом Фишера на 96.97% (для обоих коронатов) и построением деревьев принятия решения – на 90.9% для короната 18-краун-6Na⁺ и на 97.0% для короната 18-краун-6K⁺.

В этом исследовании результаты кластерного анализа методом k -средних на 100% подтверждены классификационными алгоритмами искусственной нейронной сети – многослойным персептроном и самоорганизующейся нейронной сетью Кохонена (СОК) [49, 50]. В отличие от многослойного персептрона (обучение с учителем), сеть Кохонена представляет собой соревновательную нейронную сеть с обучением без учителя.

В табл. 9, 10 приведены основные характеристики обученных классификаторов МП 4-4-3 и СОК 7-3 для короната натрия, МП 4-3-3 и СОК 5-3 для короната калия.

Проведенные исследования устойчивости коронатов натрия и калия в водно-органических растворителях методами множественной линейной регрессии и нейронных сетей показали следующее. Для подтверждения достоверности результатов решения задач многофакторного анализа (нелинейной регрессии) и классификации устойчивости коронатов необходимо использовать различные методы анализа данных. При решении задач аппроксимации констант комплексо-

Таблица 10. Основные характеристики сети Кохонена как классификатора

Сеть	Ошибка			Алгоритм обучения	Количественный состав кластеров		
	обучающая выборка, 70%	контрольная выборка, 15%	тестовая выборка, 15%		класс 1	класс 2	класс 3
СОК 7-3	18-краун-6Na ⁺			Kohonen 1000	10	12	11
	0.1215	0.1001	0.2504				
СОК 5-3	18-краун-6K ⁺			Kohonen 1000	7	10	16
	0.1025	0.1981	0.2055				

образования и прогнозирования устойчивости коронатов натрия и калия по свойствам водно-органических растворителей в равной степени применимы как классический статистический метод множественной линейной регрессии, так и нейронные сети, использующие алгоритм обучения с учителем. Метод множественной линейной регрессии по точности аппроксимации несколько уступает многослойному персептрону. Тем не менее, сильной стороной метода множественной линейной регрессии является то, что его результаты позволяют количественно оценить энергетические вклады электростатических, когезионных и донорно-акцепторных взаимодействий в возрастание устойчивости коронатов натрия и калия на основе принципа линейности свободных энергий [52].

Искусственные нейронные сети наряду с множественной линейной регрессией могут быть успешно использованы для моделирования, прогнозирования и классификации устойчивости коронатов в водно-органических растворителях. На основе построенных моделей искусственной нейронной сети и множественной линейной регрессии предсказаны константы устойчивости коронатов натрия и калия в водно-этанольных растворителях. Алгоритмы обучения многослойного персептрона с учителем и сети Кохонена, использующие алгоритм обучения без учителя, полностью подтвердили классификацию устойчивости коронатов натрия и калия, проведенную методом *k*-средних.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Artificial Neural Networks. Architectures and Applications / Ed. K. Suzuki. Chicago: University of Chicago, 2013. doi 10.5772/3409
2. Himmelblau D.M. // Korean Z. Chem. Eng. 2000. Vol. 17. N 4. P. 373. doi 10.1007/BF02706848
3. Marini F., Bucci R., Magrì A.L., Magrì A.D. // Microchem. J. 2008. Vol. 88. P. 178. doi 10.1016/j.microc.2007.11.008
4. Huang R.B., Du Q.S., Wei Y.T., Pang Z.W., Wei H., Chou K.C. // J. Theor. Biol. 2009. Vol. 256. N 3. P. 428. doi 10.1016/j.jtbi.2008.08.028
5. Бондарев Н.В. // ЖОХ. 2017. Т. 87. № 2. С. 207; Bondarev N.V. // Russ. J. Gen. Chem. 2017. Vol. 87. N 2. С. 188. doi 10.1134/S1070363217020062
6. Blanco M., Coello J., Iturriaga H., MasPOCH S., Redón M. // Anal. Chem. 1995. Vol. 67. N 24. P. 4477. doi 10.1021/ac00120a008
7. Meiler J. // J. Biomol. NMR. 2003. Vol. 26. N 1. P. 25. doi 10.1023/A:1023060720156
8. Nicelyab J.M., Haniscob T.F., Ririsb H. // J. Quant. Spectrosc. Rad. Trans. 2018. Vol. 211. P. 115. doi 10.1016/j.jqsrt.2018.03.004
9. Tetko I.V., Tanchuk V.Yu. // J. Chem. Inf. Comput. Sci. 2002. Vol. 42. N 5. P. 1136. doi 10.1021/ci025515j
10. Tetko I.V., Tanchuk V.Y., Villa A.E. // J. Chem. Inf. Comput. Sci. 2001. Vol. 41. N 5. P. 1407. doi 10.1021/ci010368v
11. Wang B., Valentine S., Plasencia M., Raghuraman S., Zhang X. // BMC Bioinformatics. 2010. Vol. 11. P. 182. doi 10.1186/1471-2105-11-182
12. Kavšek D., Bednárová A., Biro M., Kranvogel R., Vončina D.B., Beinrohr E. // Cent. Eur. J. Chem. 2013. Vol. 11. N 9. P. 1481. doi 10.2478/s11532-013-0280-x
13. Doua Y., Sunb Y., Renc Y., Rena Y. // Anal. Chim. Acta. 2005. Vol. 528. N 1. P. 55. doi 10.1016/j.aca.2004.10.050
14. Wang B., Liub G., Liuc S., Feia Q., Rena Y. // Vibr. Spectrosc. 2009. Vol. 51. N 2. P. 199. doi 10.1016/j.vibspec.2009.04.007

15. Jovanović M., Sokić D., Grabnar I., Vovk T., Prostran M., Erić S., Kuzmanovski I., Vučićević K., Miljković B. // J. Pharm. Pharm. Sci. 2015. Vol. 18. N 5. P. 856. doi 10.18433/J33031
16. de Molfetta F.A., Angelotti W.F., Romero R.A., Montanari C.A., da Silva A.B. // J. Mol. Model. 2008. Vol. 14. N 10. P. 9755. doi 10.1007/s00894-008-0332-x
17. Nandi S., Vracko M., Bagchi M.C. // Chem. Biol. Drug Des. 2007. Vol. 70. N 5. P. 424. doi 10.1111 / j.1747-0285.2007.00575.x
18. Cheng F., Vijaykumar S. // Clin. Exp. Pharmacol. 2012. Vol. 2 P. 113. doi: 10.4172/2161-1459.1000e113
19. Honório K.M., de Lima E.F., Quiles M.G., Romero R.A., Molfetta F.A., da Silva A.B. // Chem. Biol. Drug Des. 2010. Vol. 75. N 6. P. 632. doi 10.1111 / j.1747-0285.2010.00966x
20. Inci C., Ayse Y., Kürsäd U.M., Askin D., Serap C., Omca D. // J. Food Nutr. Res. 2017. Vol. 56. N 2. P. 138
21. Baha H., Dibi Z. // Sensors (Basel). 2009. Vol. 9. N 11. P. 8944. doi 10.3390/s91108944
22. Padín P.M., Peña R.M., García S., Iglesias R., Barro S., Herrero C. // Analyst. 2001. Vol. 126. N 1. P. 97. doi 10.1039/B007720H
23. Moldes O.A., Mejuto J.C., Rial-Otero R., Simal-Gandara J. // Crit. Rev. Food Sci. Nutr. 2017. Vol. 57. N 13. P. 2896. doi 10.1080/10408398.2015.1078277
24. Wine: Phenolic Composition, Classification and Health Benefits. New York: Nova Science Publishers, Inc., 2014. Ch. 10. P. 245.
25. Penza M., Cassano G. // Food Chem. 2004. Vol. 86. N 2. P. 283. doi 10.1016/j.foodchem.2003.09.027
26. Latorre M.J., Peña R., García S., Herrero C. // Analyst. 2000. Vol. 125. P. 307. doi 10.1039/A905978D
27. Cordella C.B., Militão J.S., Clément M.C., Cabrol-Bass D. // J. Agric. Food Chem. 2003. Vol. 51. N 11. P. 3234. doi 10.1021/jf021100m
28. Bos A., Bos M., van der Linden W.E. // Anal. Chim. Acta. 1992. Vol. 256. N 1. P. 133. doi 10.1016/0003-2670(92)85338-7
29. Cimpoi C., Cristea V.M., Hosu A., Sandru M., Seserman L. // Food Chem. 2011. Vol. 127. N 3. P. 1323. doi 10.1016/j.foodchem.2011.01.091
30. Angerosa F., Di Giacinto L., Vito R., Cumitini S. // J. Sci. Food Agric. 1996. Vol. 72. N 3. P. 323. doi 10.1002/(SICI)1097-0010(199611)72:3<323::AID-JSFA662>3.0.CO;2-A
31. Zhang G., Ni Y., Churchill J., Kokot S. // Talanta. 2006. Vol. 70. N 2. P. 293. doi 10.1016/j.talanta.2006.02.037
32. Cirovic D.A. // TrAC Trends Anal. Chem. 1997. Vol. 16. N 3. P.148. doi 10.1016/S0165-9936(97)00007-1
33. Meyer M., Weigelt T. // Anal. Chim. Acta. 1992. Vol. 265. N 2. P.183. doi 10.1016/0003-2670(92)85024-Z
34. Amato F., López A., Méndez E.M., Vañhara P., Hampl A., Havel J. // J. Appl. Biomed. 2013. Vol. 11. N 2. P. 47. doi 10.2478/v10136-012-0031-x
35. Maran E., Novic M., Barbieri P., Zupan J. // SAR QSAR Environ Res. 2004. Vol. 15 N 5-6. P. 469. doi 10.1080/10629360412331297461
36. Allison T.C. // J. Phys. Chem. (B). 2016. Vol. 120. N 8. P. 1854. doi 10.1021/acs.jpcc.5b09558
37. Elçiçek H., Akdoğan E., Karagöz S. // Sci. World J. 2014. Vol. 2014. P. 9. doi 10.1155/2014/194874
38. Rekha C.R., Nayar V.U., Gopchandran K.G. // Optik. 2018. Vol. 172. P. 721. doi 10.1016/j.ijleo.2018.07.090
39. Sigman M.E., Rives S.S. // J. Chem. Inf. Comput. Sci. 1994. Vol. 34. N 3. P. 617. doi 10.1021/ci00019a021
40. Myint K.Z., Xie X-Q. // Methods Mol. Biol. 2015. N 1260. P. 149. doi 10.1007/978-1-4939-2239-0_9
41. Wei J.N., Duvenaud D., Aspuru-Guzik A. // ACS Cent. Sci. 2016. Vol. 2. N 10. P. 725. doi 10.1021/acscentsci.6b00219
42. DiRusso S.M., Sullivan T., Holly C., Cuff S.N., Savino J. // J. Trauma. 2000. Vol. 49. N 2. P. 212. doi 10.1097/00005373-200008000-00006
43. Бондарев Н.В. Термодинамика равновесий. Эффекты среды и нейросетевой анализ. Saarbrücken: LAP LAMBERT Academic Publishing, 2012. 380 с.
44. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. СПб: Питер, 2003. 686 с.
45. Боровиков В.П. Популярное введение в современный анализ данных в системе STATISTICA. Учебное пособие для вузов. М.: Горячая линия – Телеком, 2013. 288 с.
46. Хайкин С. Нейронные сети. М.: Вильямс, 2006. 1104 с.
47. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. М.: Мир, 1985. 509 с.
48. Бондарев Н.В. // ЖОХ. 2019. Т. 89. № 2. С. 288. doi 10.1134/S0044460X19020197. Bondarev N.V. // Russ. J. Gen. Chem. 2019. Vol. 89. No. 2. P. 281. doi 10.1134/S1070363219020191
49. Кохонен Т. Самоорганизующиеся карты. М.: Лаборатория знаний, 2017. 660 с.
50. Холін Ю.В., Пушкарьова Я.М., Пантелеймонов А.В., Некос А.Н. Хемометричні методи в розв'язанні задач якісного хімічного аналізу та класифікації фізико-хімічних даних. Х.: ХНУ імені В.Н. Каразіна, 2016. 184 с.
51. de Boer, P.-T., Kroese, D., Mannor, S., Rubinstein, R.Y. // Ann. Oper. Res. 2005. Vol. 134. N 1. P. 19. doi 10.1007/s10479-005-5724-z
52. Шмид Р., Сапунов В.Н. Неформальная кинетика. В поисках путей химических реакций. М.: Мир, 1985. 264 с.

Artificial Neural Network and Multiple Linear Regression for Prediction and Classification of Sustainability of Sodium and Potassium Coronates

N. V. Bondarev*

V.N. Karazin Kharkiv National University, pl. Svobody 4, Kharkiv, 61022 Ukraine

**e-mail: bondarev_n@rambler.ru*

Received January 23, 2019; revised January 23, 2019; accepted February 10, 2019

Models of multiple linear regression and a multilayer artificial neural network were developed for modeling and predicting the stability constants of sodium and potassium coronates using the properties of water-organic solvents water–methanol, water–propan-2-ol, water–acetonitrile, and water–acetone. The predicted values of the stability constants of coronates in water-ethanol solvents are given and the results of the models of multiple linear regression and an artificial neural network were compared. Based on the models of multiple linear regression and the principle of linearity of free energies, a quantitative assessment of the contributions of electrostatic, cohesive and electron-donating interactions to the increase in the stability of coronates was carried out. Neural network models were developed to classify the stability of sodium and potassium coronates based on unsupervised learning algorithms (multilayer perceptrons) and with a teacher (Kohonen networks). It was shown that neural network classifiers fully confirmed the classification of the stability of coronates, conducted by the *k*-means exploration method.

Keywords: multiple regression, multilayer perceptron, Kohonen network, prediction, neural network classifier