

УДК 544.354.081.7:004.021

МЕТОДЫ РАЗВЕДОЧНОГО АНАЛИЗА И ХЕМОИНФОРМАТИКИ В ТЕРМОДИНАМИКЕ РАВНОВЕСИЙ

© 2010 С. Н. Бондарев, Н. В. Бондарев

Впервые применены разведочные методы – факторный анализ, кластерный анализ, дискриминантный анализ, канонический анализ, построение деревьев классификации и методы химической информатики – множественная линейная регрессия, аппарат нейронных сетей персептронного типа для классификации и прогнозирования констант диссоциации уксусной кислоты и устойчивости коронатов калия на основе физико-химических свойств водно-органических растворителей. Построены решающие правила классификации и трехслойные персептроны, позволяющие прогнозировать константы химических равновесий диссоциации и комплексообразования. Показана перспективность методологии сочетания нестатистических методов, сольватационно-термодинамического подхода и нейросетевых технологий для выявления закономерностей связи термодинамических характеристик химических равновесий и сольватации реагентов, классификации и прогнозирования силы карбоновых кислот и устойчивости коронатов в разных растворителях.

Ключевые слова: разведочные методы, хемоинформатика, искусственная нейронная сеть, персептрон, множественная регрессия, массив данных, константа равновесия, энергия Гиббса, сольватация.

ВВЕДЕНИЕ

Наличие в физикохимии растворов больших массивов экспериментальных данных обуславливает применение разведочных математических методов анализа, позволяющих перед проведением углубленного многомерного статистического анализа структурировать информацию, находить объективные закономерности в больших объемах данных, выявлять прогностические возможности обобщений и правил. В то же время, проведение глубокого анализа невозможно без использования средств современной информатики – "науки о принципиально новой человеко-машинной технологии расширенного воспроизводства качественно нового знания" [1]. Вследствие этого на стыке химии и информатики возникает и быстро оформляется в самостоятельную научную дисциплину *хемоинформатика* [2-4], методы которой начинают активно внедряться во все области химии. Развитию хемоинформатики в значительной мере способствует наличие обоснованной методологии и реализующего ее программного обеспечения, которые позволяют химику на основе обработки экспериментальных данных осуществлять прогнозирование самых разнообразных свойств химических соединений и процессов [5-7]. При этом на первый план выходят методы нелинейного моделирования, в частности технологии искусственных нейронных сетей (ИНС) [8].

Новизна данной работы заключается в применении многомерных разведочных методов анализа к термодинамике равновесий диссоциации и комплексообразования для построения решающих правил классификации и прогнозирования силы уксусной кислоты и устойчивости коронатов калия в водно-органических растворителях, а также в демонстрации согласованности традиционных подходов (регрессионно-корреляционный, сольватационно-термодинамический) с современными нейросетевыми методами анализа и прогнозирования термодинамических свойств (изменений свободной энергии Гиббса, энтальпии, энтропии и констант) химических равновесий.

Сольватационно-термодинамический метод [9] дает возможность выявить вклады эффектов среды (сольватации реагентов) в изменение термодинамических параметров химических равновесий. Особенностью регрессионно-корреляционного подхода является принцип линейности свободных энергий (ЛСЭ), позволяющий вскрыть причины (электростатические, химические взаимодействия) влияния растворителя на термодинамику химических реакций [10]. Искусственные нейронные сети широко применяются в органической, аналитической, физической и биологической областях химии для обработки массивов экспериментальных данных с целью построения прогностических моделей [11-14].

СПЕЦИФИКА РАЗВЕДОЧНЫХ И НЕЙРОСЕТЕВЫХ МЕТОДОВ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ДАННЫХ

Многомерные разведочные методы анализа. С помощью факторного анализа проверялись все возможные варианты взаимосвязей между переменными для их редуцирования. С математической точки зрения факторная модель имеет вид [15]

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \dots + A_{im}F_m + V_iU_i,$$

где X_i – i -я нормированная переменная; A_{im} – нормированный коэффициент множественной регрессии переменной i по общему фактору m ; F_i – общий фактор; V_i – нормированный коэффициент регрессии переменной i по характерному фактору; U_i – характерный фактор для переменной i ; m – число общих факторов.

Характерные факторы не коррелируют между собой и с общими факторами. Общие факторы можно выразить линейными комбинациями наблюдаемых переменных:

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \dots + W_{ik}X_k,$$

где F_i – оценка i -го фактора; W_i – весовой коэффициент или коэффициент значения фактора; k – число переменных. Веса подбирались так, чтобы первый коэффициент значения фактора объяснял наибольшую долю полной дисперсии. Затем отбирался второй набор весов так, чтобы второй фактор вносил наибольший вклад в остаточную дисперсию при условии, что он не коррелирует с первым фактором.

Задача кластерного анализа [16] заключалась в разделении констант равновесий диссоциации или комплексообразования на однородные группы, характеризующие силу уксусной кислоты (умеренно слабая, слабая, очень слабая) или устойчивость коронатов калия (слабоустойчивые, умеренно устойчивые, устойчивые), исходя из рассматриваемого набора переменных (свойств водно-органических растворителей). В работе реализованы иерархические методы кластерного анализа: древовидная кластеризация (агломеративный метод) и дивизивный метод k -средних.

При дискриминантном анализе [17] зависимая переменная (константа диссоциации кислоты или комплексообразования) имела статус категориальной, а предикторы (независимые переменные) были интервальными. Модель дискриминантного анализа (дискриминантная функция) имеет следующий вид:

$$D = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k,$$

где D – дискриминантный показатель (дискриминант), b – дискриминантный коэффициент или вес, X – предиктор или независимая переменная.

С помощью канонического анализа оценивалась степень связи между дискриминантными показателями и группами (кластерами).

Таким образом, кластерный, дискриминантный и канонический анализ наиболее полно отражают особенности многомерного разведочного анализа в классификации силы кислоты и устойчивости коронатов, а факторный – в исследовании взаимосвязи между параметрами.

Основное назначение построения деревьев классификации (деревьев решений) заключалось в подведении итогов разведочного анализа, в структурировании данных и построении решающих правил классификации (принятия решений), обладающих высоким прогностическим потенциалом [18,19].

Искусственные нейронные сети. Для решения задачи регрессии методами нейросетевого моделирования в работе проанализированы сети следующих типов: линейная сеть (ЛС), радиальная базисная функция (РБФ) и многослойный персептрон (МП) [8,20].

Линейная модель, по сути, ничем не отличается от обычной линейной регрессии, но на языке нейронных сетей представляется сетью без промежуточных слоев. ЛС в выходном слое содержит только линейные элементы, т.е. нейроны с линейной функцией активации [20].

Общий вид радиально-базисной функции [8]: $f(x) = \varphi\left(\frac{x^2}{\sigma^2}\right)$, например, $f(x) = e^{-\frac{x^2}{\sigma^2}}$, где

x – вектор входных сигналов нейрона, σ – ширина окна функции, $\varphi(y)$ – убывающая функция (чаще всего, равная нулю вне некоторого отрезка). Радиально-базисная сеть характеризуется тремя особенностями: единственным скрытым слоем; нелинейную активационную функцию

имеют только нейроны скрытого слоя; веса синаптических связей входного и скрытого слоев равны единице.

Теоретическое обоснование нейросетевого моделирования с помощью многослойных персептронов (МП) базируется на теореме А.Н.Колмогорова, доказавшего [21], что любую непрерывную многомерную функцию на единичном отрезке $[0;1]$ можно представить в виде конечного числа одномерных функций:

$$f(x_1, x_2, \dots, x_n) = \sum_{p=1}^{2n+1} g \left(\sum_{i=1}^n \lambda_i \varphi_p(x_i) \right)$$

где g и φ_p непрерывные и одномерные функции, $\lambda_i = \text{const}$.

Вычисления, проводимые МП, описываются формулой

$$F(x) = \sum_{k=1}^H V_k f(W_k x + b_k),$$

здесь H – число нейронов скрытого слоя, f – нелинейная функция преобразования входного сигнала нейрона в выходной, W и V – весовые коэффициенты связей от входа к скрытому слою и от скрытого слоя к выходу, соответственно, b – аддитивная составляющая входного сигнала (порог возбуждения нейрона).

Цель настоящей работы: на основе многомерных разведочных методов анализа свободных энергий Гиббса (констант) химических равновесий и свойств водно-органических растворителей построить решающие правила классификации силы уксусной кислоты и устойчивости коронатов калия; построить регрессионные линейные и нейросетевые модели персептронного типа для прогнозирования силы уксусной кислоты и устойчивости коронатов калия в разных растворителях; сопоставить результаты сольватационно-термодинамического и корреляционно-регрессионного анализа термодинамики диссоциации уксусной кислоты и комплексообразования эфира 18-краун-6 (L) с катионами калия в водно-органических растворителях.

Анализируемые данные: зависимые переменные – аквамолярные [22] стандартные энергии Гиббса диссоциации ($\Delta G_{\text{д,НАс}}^{\circ}$) уксусной кислоты [23-25] и комплексообразования катионов калия с эфиром 18-краун-6 ($\Delta G_{\text{LK}^+}^{\circ}$) [26-35]; независимые переменные – физико-химические свойства водно-органических растворителей (вода-метанол, вода-этанол, вода-пропан-2-ол, вода-ацетонитрил, вода-ацетон, вода-диоксан) – диэлектрическая проницаемость, электроноакцепторные параметры Димрота-Райхардта, электронодонорные параметры Камлета-Тафта и плотность энергии когезии [36-39].

МНОГОМЕРНЫЙ РАЗВЕДОЧНЫЙ АНАЛИЗ

Редукция данных проведена с помощью факторного анализа, целесообразность использования которого подтверждена критерием сферичности Бартлетта и критерием адекватности выборки Кайзера-Мейера-Олкина (КМО) [15,16]. Факторы выделены методом главных компонент. Число значимых факторов выбрано равным двум по критерию каменистой осыпи. Факторные нагрузки были найдены методом вращения варимакс [40]. Процедура редукции данных позволила выделить два значимых фактора, влияющих на силу уксусной кислоты и устойчивость коронатов калия в водно-органических растворителях.

Решением задачи *кластерного анализа* (древовидная классификация по методу Варда, метод k -средних) явилось разбиение массива данных по константам диссоциации уксусной кислоты и устойчивости коронатов калия на три класса, удовлетворяющее критерию оптимальности [40].

На основе дискриминантного и канонического анализа установлено, что наибольший вклад в дискриминацию силы уксусной кислоты вносит переменная $1/\varepsilon^N$, а в различии устойчивости коронатов калия наибольшую роль играют плотность энергии когезии δ_N^2 и опять-таки обратная величина диэлектрической проницаемости водно-органических растворителей.

Для анализа структуры данных были построены деревья классификации для стандартной энергии Гиббса диссоциации уксусной кислоты (табл. 1) и устойчивости коронатов калия (табл. 2) в водно-органических растворителях.

Получено решающее правило, позволяющее классифицировать уксусную кислоту как умеренно слабую, слабую или очень слабую по значениям диэлектрической проницаемости водно-органических растворителей.

Таблица 1. Структура дерева классификации силы уксусной кислоты по значениям обратной величины диэлектрической проницаемости растворителей

Структура: дочерние вершины, наблюдаемые классы, предсказанные классы, условия ветвления								
Вершина	Левая вершина	Правая вершина	Класс 1	Класс 2	Класс 3	Предсказанный класс	Ветвление по постоянной	Ветвление по переменной
1	2	3	9	11	13	3	-1.665	$1/\varepsilon^N$
2			0	0	13	3		
3	4	5	9	11	0	2	-2.696	$1/\varepsilon^N$
4			1	10	0	2		
5			8	1	0	1		

Если неравенство $1/\varepsilon^N \leq 1.665$ выполняется, то константы диссоциации классифицируют кислоту как умеренно слабую (класс 3, рК: 4.76 – 5.71). Если $1/\varepsilon^N > 1.665$, константы диссоциации определяют кислоту как слабую или очень слабую. Затем проверяется справедливость неравенства $1/\varepsilon^N \leq 2.696$. Если неравенство выполняется, константы характеризуют кислоту как слабую (класс 2, рК: 5.94 – 7.58), а если $1/\varepsilon^N > 2.696$, кислота классифицируется как очень слабая (класс 1, рК: 7.02 – 9.91).

Таблица 2. Структура дерева классификации устойчивости короната калия по значениям обратной величины диэлектрической проницаемости и параметра Камлета-Тафта разных растворителей

Структура: дочерние вершины, наблюдаемые классы, предсказанные классы, условия ветвления								
Вершина	Левая вершина	Правая вершина	Класс 1	Класс 2	Класс 3	Предсказанные классы	Ветвление по постоянной	Ветвление по переменной
1	2	3	7	10	16	3	-1.672	$1/\varepsilon^N$
2			0	0	15	3		
3	4	5	7	10	1	2	-0.635	$B_{КТ}$
4			0	10	0	2		
5			7	0	1	1		

Таблица 3. Подтверждение выполнимости решающего правила классификации силы уксусной кислоты в водно-органических растворителях

мол. доля DO	$1/\varepsilon^N$	рК _{эсп}	Правило, класс (кластер), интервал изменений рК			
			Вода – диоксан (DO) – уксусная кислота			
0	1	4.76	$1/\varepsilon^N \leq 1.665$		3 4.76 – 5.71	
0.1	1.645	5.70	$1/\varepsilon^N \leq 1.665$		3 4.76 – 5.71	
0.2	2.625	6.67	$1.665 < 1/\varepsilon^N < 2.696$		2 5.94 – 7.58	
0.3	3.848	7.71	$1/\varepsilon^N > 2.696$		1 7.02 – 9.91	
0.4	5.326	8.90	$1/\varepsilon^N > 2.696$		1 7.02 – 9.91	
0.5	9.229	10.31	$1/\varepsilon^N > 2.696$		1 7.02 – 9.91	
мол. доля DMSO	Вода – диметилсульфоксид (DMSO) – уксусная кислота					
	0	1	4.76	$1/\varepsilon^N \leq 1.665$		3 4.76 – 5.71
	0.1	1.025	5.17	$1/\varepsilon^N \leq 1.665$		3 4.76 – 5.71
	0.2	1.047	5.79	$1/\varepsilon^N \leq 1.665$		3 4.76 – 5.71

Решающее правило прогнозирования устойчивости комплексов эфира 18-краун-6 с катионами K^+ позволяет классифицировать устойчивость коронатов калия по значениям диэлектрической проницаемости и параметру Камлета-Тафта водно-органических растворителей.

При выполнении неравенства $1/\varepsilon^N \leq 1.672$ константы классифицируют коронат калия как комплекс слабоустойчивый (класс 3, lg K: 2.04 – 3.74). Если $1/\varepsilon^N > 1.672$, константы определяют коронат калия как умеренно устойчивый или устойчивый. На втором этапе для констант

устойчивости проверяется справедливость неравенства $B_{КТ} \leq 0.635$. Если неравенство выполняется, константы характеризуют коронат калия как устойчивый (класс 2, $\lg K$: 3.80 – 5.84). Если $B_{КТ} > 0.635$, то константы классифицируют коронат калия как умеренно устойчивый комплекс (класс 1, $\lg K$: 3.18 – 3.49).

Прогностический потенциал полученных решающих правил проверен на независимых экспериментальных данных – константах диссоциации уксусной кислоты в водно-диоксановых и водно-диметилсульфоксидных растворителях (табл. 3), а также на константах устойчивости коронатов калия в водно-диоксановых и водно-ацетоновых растворителях (табл. 4).

Таблица 4. Подтверждение выполнимости решающего правила классификации устойчивости коронатов калия в водно-органических растворителях

мол. доля DO	$1/\varepsilon^N$	$B_{КТ}$	$\lg K_{\text{эсп}}$	Правило, класс (кластер), интервал изменений $\lg K$		
				Вода – диоксан (DO) – $18C6K^+$		
0	1	0.19	2.04	$1/\varepsilon^N \leq 1.672$	3	2.04 – 3.74
0.1	1.645	0.36	2.67	$1/\varepsilon^N \leq 1.672$	3	2.04 – 3.74
0.2	2.625	0.45	3.32	$1/\varepsilon^N > 1.672$ $B_{КТ} \leq 0.63$	3, 2	2.04 – 5.84
0.3	3.848	0.48	3.83	$1/\varepsilon^N > 1.751$ $B_{КТ} \leq 0.63$	2	3.80 – 5.84
0.4	5.326	0.50	4.02	$1/\varepsilon^N > 1.751$ $B_{КТ} \leq 0.63$	2	3.80 – 5.84
мол. доля А	Вода – ацетон (А) – $18C6K^+$					
	0	1.000	0.19	2.04	$1/\varepsilon^N \leq 1.672$	3
0.1	1.237	0.40	2.53	$1/\varepsilon^N \leq 1.672$	3	2.04 – 3.74
0.2	1.514	0.49	2.99	$1/\varepsilon^N \leq 1.672$	3	2.04 – 3.74
0.3	1.820	0.52	3.41	$1/\varepsilon^N > 1.672$ $B_{КТ} \leq 0.63$	3, 2	2.04 – 5.84
0.4	2.143	0.52	3.80	$1/\varepsilon^N > 1.672$ $B_{КТ} \leq 0.63$	2	3.80 – 5.84
0.5	2.473	0.53	4.13	$1/\varepsilon^N > 1.672$ $B_{КТ} \leq 0.63$	2	3.80 – 5.84

МНОЖЕСТВЕННЫЙ ЛИНЕЙНЫЙ АНАЛИЗ

Изменение стандартной энергии Гиббса диссоциации уксусной кислоты ($\Delta G^\circ_{d, \text{HAc}}$). Для проведения анализа использованы данные по энергиям Гиббса диссоциации уксусной кислоты в растворителях вода-метанол, вода-этанол, вода-пропан-2-ол при 298.15K с содержанием неводного компонента до 0.7 мол. доли (до области пересольватации [41], после которой изменяется состав сольватных оболочек ионов и молекул кислоты и проявляется нелинейность зависимости $\Delta G^\circ_{d, \text{HAc}}$ от свойств растворителей).

Значимые дескрипторы и наличие мультиколлинеарности выявлены с помощью корреляционной матрицы и пошаговым включением параметров в уравнение регрессии. Критерием включения дескриптора в уравнение регрессии был выбран уровень значимости ($p < 0.05$). Коэффициенты уравнения множественной регрессии оценены стандартным методом наименьших квадратов (МНК).

Статистически значимыми факторами, определяющими зависимость энергии Гиббса диссоциации уксусной кислоты от свойств растворителей, выявлены диэлектрическая проницаемость и плотность энергии когезии.

В табл. 5 приведены результаты статистического анализа для уравнения регрессии

$$\Delta G^\circ_{d(\text{HAc})} = (37.3 \pm 2.8) + (3.6 \pm 0.5) \cdot \varepsilon_N^{-1} - (13.3 \pm 2.7) \cdot \delta_N^2 \quad (1)$$

При использовании уравнения (1) для оценки энергии Гиббса диссоциации уксусной кислоты в водно-диоксановых и водно-диметилсульфоксидных растворителях обнаружена удовлетворительная прогностическая способность двухпараметрической линейной модели для водно-органических растворителей с малым содержанием диоксана либо диметилсульфоксида табл. 6.

Согласно уравнению (1), с уменьшением диэлектрической проницаемости и уменьшением плотности энергии когезии при замене воды на водно-органические растворители сила уксусной кислоты понижается. Это означает, что: а) уменьшение диэлектрической проницаемости растворителя сопровождается уменьшением энергии Гиббса сольватации ионов, следствием чего является сдвиг равновесия в сторону недиссоциированных молекул кислоты; б) при

уменьшении плотности энергии когезии водно-органические растворители в большей мере стабилизируют недиссоциированную форму уксусной кислоты (рис. 1 в качестве примера).

Таблица 5. Итоги регрессии* для зависимой переменной $\Delta G_d^0(\text{HAc})$

$R = 0.99$; $R^2 = 0.98$; $F(2,19) = 454.02$; Стандартная ошибка оценки: 0.64

	<i>Beta</i>	Стандартная ошибка <i>Beta</i>	B_k	Стандартная ошибка B_k	$t(19)$	p-уровень
			37.3	1.4	27.2	0.00
$1/\varepsilon_N$	0.64	0.04	3.6	0.2	14.7	0.00
δ_N^2	-0.44	0.04	-13.3	1.3	-10.0	0.00

*Примечание: *Beta* – коэффициенты уравнения $u_i = \sum_{k=1}^p \text{Beta}_k z_i^{(k)}$, в котором $u_i = (y_i - \bar{y}) / S_y$ и $z_i^{(k)} = (x_i^{(k)} - \bar{x}^{(k)}) / S_{x^{(k)}}$ – стандартизованные переменные, y_i – экспериментальные значения $\Delta G_d^0(\text{HAc})$; \bar{y} – среднее значение $\Delta G_d^0(\text{HAc})$, $x_i^{(k)}$ – значения независимой переменной (ε_N^{-1} или δ_N^2), $\bar{x}^{(k)}$ – среднее значение независимой переменной; S_y и $S_{x^{(k)}}$ – стандартные ошибки зависимой $\Delta G_d^0(\text{HAc})$ и независимых переменных; R – коэффициент корреляции; R^2 – коэффициент детерминации; $t(19)$ – наблюдаемый коэффициент Стьюдента $t = B_k / S_{B_k}$ для числа степеней свободы 19.

Таблица 6. Сопоставление наблюдаемых значений энергии Гиббса диссоциации уксусной кислоты с предсказанными по уравнению (1)

мол. доля DO или DMSO	Свойства растворителя				$\Delta G_{d,\text{HAc}}^0$ наблюдаемые	$\Delta G_{d,\text{HAc}}^0$ предсказанные	Остатки
	$1/\varepsilon^N$	E_T^N	B_{KT}	δ_N^2			
Вода-диоксан							
0	1	1	0.19	1	27.15	27.62	-0.47
0.1	1.645	0.77	0.36	0.917	32.56	31.06	1.50
0.2	2.625	0.69	0.45	0.834	38.06	35.71	2.35
0.3	3.848	0.62	0.48	0.751	44.02	41.24	2.78
0.4	5.326	0.60	0.50	0.668	50.83	47.70	3.13
0.5	9.229	0.54	0.51	0.584	58.88	62.94	-4.06
Вода-диметилсульфоксид							
0	1	1	0.19	1	27.15	27.62	-0.47
0.1	1.025	0.85	0.35	0.931	29.51	28.63	0.88
0.2	1.047	0.75	0.45	0.862	33.03	29.63	3.40

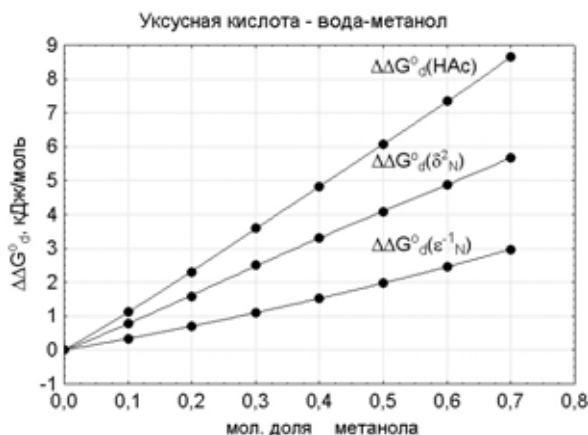


Рис. 1. Влияние плотности энергии когезии и диэлектрической проницаемости растворителей вода-метанол на уменьшение силы уксусной кислоты $\Delta \Delta G_d^0(\text{HAc}) = \Delta G_d^{0,w-s}(\text{HAc}) - \Delta G_d^{0,w}(\text{HAc})$.

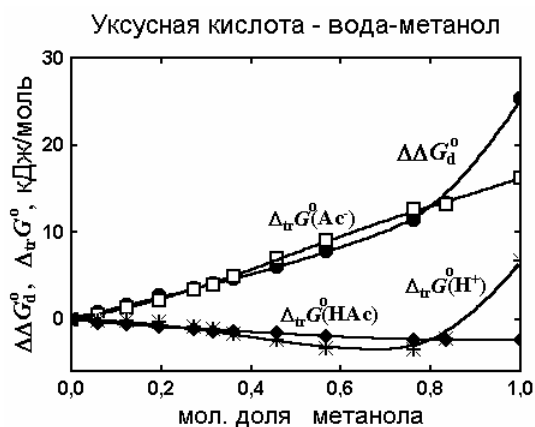


Рис. 2. Влияние энергии Гиббса пересольватации ионов $\Delta_{tr}G^0(\text{H}^+)$, $\Delta_{tr}G^0(\text{Ac}^-)$ и молекул $\Delta_{tr}G^0(\text{HAc})$ на изменение энергии Гиббса диссоциации уксусной кислоты $\Delta \Delta G_d^0(\text{HAc})$ при замене воды на водно-метанольные растворители.

Сопоставление результатов линейного многофакторного регрессионного анализа (рис. 1) и результатов сольватационно-термодинамического анализа (рис. 2 [42] в качестве примера) позволяет говорить об их адекватности. Уменьшение силы уксусной кислоты в водно-метанольных растворителях по сравнению с водой обусловлено возрастанием сольватации недиссоциированных молекул кислоты и уменьшением сольватации ацетат-ионов (рис. 2).

Изменение стандартной энергии Гиббса устойчивости коронатов калия ΔG_{LK}^{o+} . Для анализа использованы данные по энергиям Гиббса образования комплексов LK^+ в растворителях вода-метанол, вода-пропан-2-ол, вода-ацетонитрил, вода-диоксан и вода-ацетон при 298.15 К с содержанием неводного компонента до 0.5 мол. доли.

Отбор значимых независимых дескрипторов в регрессионное уравнение проведен методом пошагового включения/исключения [16]. В качестве критерия полезности переменных был выбран уровень значимости $\text{sig } F \leq 0.05$ для включения и $\text{sig } F \geq 0.1$ для исключения переменных из регрессионного уравнения. Значимость включения устанавливалась меньше значимости исключения во избежание закливания процесса включения/исключения.

Результаты проведенного статистического анализа (табл. 7,8) свидетельствуют о значимости дескрипторов δ_N^2 и $1/\epsilon^N$. Приведенные в табл. 7 данные показывают, как изменялись характеристики регрессионного анализа при поэтапном включении новых предикторов в уравнение регрессии. С увеличением числа предикторов R , R^2 и исправленная величина R_a^2 возрастают, в то время как значение стандартной ошибки убывает. Если для первой модели 93.3 % ($R^2 = 0.933$) дисперсии энергии Гиббса комплексообразования объясняется влиянием только плотности энергии когезии, то для второй модели совокупное воздействие диэлектрической проницаемости и плотности энергии когезии определяет уже 96.7 % дисперсии ΔG_{LK}^{o+} .

Полученное значение критерия Дарбина-Уотсона (табл. 7) $d = 1.07$ (коэффициент автокорреляции остатков $\rho = 0.463$) свидетельствует о тенденции к положительной автокорреляции в остатках ($d_L = 1.37$, $d_U = 1.59$), что подразумевает дальнейшее пополнение массива данных для корректировки модели. Трудность решения этой проблемы обусловлена тем, что общий вид зависимости неучтенных в модели факторов и изменений энергии Гиббса комплексообразования неизвестен, а это в значительной мере осложняет применение стандартного аппарата математической статистики. Одно из преимуществ нейронных сетей состоит в том, что при нейросетевом моделировании не делается предположений об истинной форме этих связей.

Таблица 7. Статистические характеристики [43] регрессионных моделей

Модель	R	R ²	R ² _a скорректированный	Стандартная ошибка регрессии, S	Изменение статистик					Критерий Дарбина-Уотсона, d
					R ² _{ch} изменение	F _{ch} изменение	df1	df2	sig F изменение	
1 ^a	0.97	0.94	0.93	1.02	0.94	415.96	1	29	0.00	
2 ^b	0.98	0.96	0.97	0.73	0.03	27.58	1	28	0.00	1.07

Примечание: ^a $\Delta G_{LK}^{o+} = B_0 + B_1 \delta_N^2$; ^b $\Delta G_{LK}^{o+} = B_0 + B_1 \delta_N^2 + B_2(1/\epsilon_N)$; R – коэффициент множественной корреляции; $R^2 = (1 - SS_{res} / SS_t)$ – коэффициент детерминации (смещенная оценка);

$R_a^2 = 1 - \frac{SS_{reg} / (N - p - 1)}{SS_t / (N - 1)}$ – скорректированная оценка коэффициента детерминации;

$SS_t = \sum (y_i - \bar{y})^2$ – общая сумма квадратов отклонений; $SS_{res} = \sum (y_i - \hat{y}_i)^2$ – сумма квадратов регрессионных отклонений; $SS_{reg} = \sum (\hat{y}_i - \bar{y})^2$ – разброс по линии регрессии;

$S^2 = (\sum (y_i - \hat{y}_i) / (N - p - 1))$ – дисперсия случайной ошибки (аппроксимации), y_i – экспериментальные значения ΔG_{LK}^{o+} , \bar{y} – среднее значение ΔG_{LK}^{o+} , \hat{y}_i – рассчитанные по уравнению регрессии зна-

чения $\Delta G_{LK}^{\circ+}$; $R_{ch}^2 = (R^2 - R_{(k)}^2)$ – уменьшение объясненной дисперсии; $R_{(k)}^2$ – коэффициент детерминации, полученный при исключении зависимой переменной x^k ; $F_{ch} = R_{ch}^2(N - p - 1) / (1 - R^2)$ – значимость включения в уравнение регрессии переменной x^k ; df – число степеней свободы;

$F = \frac{SS_{reg} / p}{SS_{res} / (N - p - 1)}$ – распределение (критерий) Фишера; N – число экспериментальных значений

$\Delta G_{LK}^{\circ+}$; p – число параметров уравнения регрессии без учета свободного члена; $\text{sig } F$ – наблюдаемая значимость статистики Фишера (уровень значимости).

На основе полученного уравнения множественной линейной регрессии (табл. 8)

$$\Delta G_{LK}^{\circ+} = (-35.5 \pm 2.6) + (24.5 \pm 2.6) \cdot \delta_N^2 - (0.92 \pm 0.36) \cdot (1/\varepsilon_N) \quad (2)$$

проведена количественная оценка влияния диэлектрической проницаемости и плотности энергии когезии на устойчивость коронатов калия в водно-органических растворителях.

Таблица 8. Итоги регрессии для двухпараметрической $(1/\varepsilon_N, \delta_N^2)$ зависимой переменной $\Delta G_{LK}^{\circ+}$

$R = 0.98$; $R^2 = 0.96$; $F(2,28) = 411.53$; $(F_{кр}(2,28) = 3.34)$; Стандартная ошибка оценки: 0.74						
$N = 31$	Бета	Стандартная ошибка Бета	B	Стандартная ошибка B	$t(28)$	$\text{sig } F$
Свободный член			-35.5	1.27	-27.81	0.00
$1/\varepsilon_N$	-0.23	0.043	-0.92	0.18	-5.24	0.00
δ_N^2	0.83	0.043	24.5	1.27	19.23	0.00

Уменьшение диэлектрической проницаемости и плотности энергии когезии водно-органических растворителей по сравнению с водой способствует увеличению устойчивости монокоронатов калия, так как $\Delta \Delta G_{LK}^{\circ+}(1/\varepsilon_N) < 0$ и $\Delta \Delta G_{LK}^{\circ+}(\delta_N^2) < 0$. Это можно объяснить тем, что уменьшение диэлектрической проницаемости в большей мере способствует уменьшению энергии Гиббса сольватации катионов K^+ , чем уменьшению энергии Гиббса сольватации комплексных ионов LK^+ . В то же время образованию полостей в водно-органических растворителях содействует уменьшение плотности энергии когезии, вследствие чего повышается стабилизация растворенных частиц водно-органическим растворителем. Таким образом, влияние когезионных взаимодействий в большей мере проявляется в стабилизации комплексных катионов LK^+ , чем катионов K^+ и лиганда 18C6, что в итоге приводит к увеличению сольватации комплексов $\Delta \Delta G_{LK}^{\circ+}(\delta_N^2) < 0$.

Выводы, вытекающие из корреляционно-регрессионного анализа, подтверждают результаты сольватационно-термодинамического анализа влияния свойств водно-органических растворителей на устойчивость коронатов калия [26]. Возрастание устойчивости коронатов калия в водно-метанольных растворителях (до $X_{MeOH} = 0.5$) обусловлено антагонистическим эффектом пересольватации реагентов [9] – суммарное уменьшение энергии Гиббса сольватации лиганда и катиона калия превышает энергию Гиббса пересольватации комплексного катиона. В отличие от водно-метанольных растворителей, в водно-пропан-2-ольных растворителях наблюдается незначительное возрастание устойчивости сольватов LK^+ .

НЕЙРОСЕТЕВОЙ ПОДХОД К АНАЛИЗУ И ПРОГНОЗИРОВАНИЮ СИЛЫ УКСУСНОЙ КИСЛОТЫ И УСТОЙЧИВОСТИ КОРОНАТОВ КАЛИЯ

Изменение стандартной энергии Гиббса диссоциации уксусной кислоты ($\Delta G_{d,HA}^{\circ}$). Предварительно были проанализированы 1000 сетей (линейная, радиальная базисная функция (число скрытых элементов $min = 1$, $max = 8$); трехслойный персептрон (число скрытых элементов $min = 1$, $max = 10$) и выбран перспективный тип сети и вариант архитектуры – это трехслойный персептрон с пятью скрытыми нейронами МП 4:4-5-1:1 (табл. 9).

В рассматриваемой регрессионной задаче линейная нейросетевая модель характеризуется самой низкой производительностью (табл. 9) – значение стандартного отклонения на контрольной выборке равно 0.424, для модели радиальной базисной функции (РБФ) это значение составляет 0.253. Сеть с архитектурой МП 4:4-8-1:1 имеет несколько лучшие статистические характеристики (контрольная производительность равна 0.107) по сравнению с МП 4:4-5-1:1

(контрольная производительность равна 0.114), однако, в случае небольших массивов данных, предпочтение отдается менее сложным архитектурам сетей.

Дальнейшая корректировка сети была проведена с помощью алгоритмов быстрого распространения (100 эпох на первом этапе; эпохе в итерационном обучении нейронной сети соответствует один проход по всему обучающему множеству с последующей проверкой на контрольном множестве) и Левенберга-Маркара (500 эпох на втором этапе), который считается одним из лучших алгоритмов нелинейной оптимизации [44].

В работе также были проанализированы перцептроны МП 4:4-3-1:1 и МП 4:4-4-1:1, содержащие в скрытом слое соответственно три и четыре нейрона. Однако, несмотря на достаточно высокую способность к аппроксимации данных, эти сети обладают меньшей способностью к обобщению (прогнозированию).

Для обучения сетей все множество энергий Гиббса диссоциации уксусной кислоты (наблюдений) было разбито на три выборки. По умолчанию осуществлялось случайное распределение наблюдений между выборками во избежание переобучения сети и для гарантирования качественного обобщения (прогнозирования). Первая выборка (*обучающая* – 50% наблюдений) использовалась для обучения сети; вторая (*контрольная* – 25% наблюдений) – для кросс-валидации алгоритма обучения во время его работы, и третья (*тестовая* – 25% наблюдений) – для окончательного независимого тестирования. Обучение проведено со скоростью 0.01.

Таблица 9. Статистические характеристики искусственных нейронных сетей разного типа

	Архитектура, коэффициент корреляции Пирсона	Производительность обучения	Контрольная производительность	Тестовая производительность	Ошибка обучения	Контрольная ошибка	Тестовая ошибка	Обучение, элементы*
1	Линейная 2:2-1:1, 0.892	0.417	0.424	0.465	0.139	0.082	0.198	ПО
2	РБФ 4:4-7-1:1 0.987	0.120	0.253	0.222	0.013	0.010	0.029	КС, КБ, ПО
3	МП 4:4-5-1:1, 0.995	0.090	0.114	0.119	0.030	0.014	0.048	ОР100, СГ20, СГ31b
4	МП 4:4-8-1:1, 0.997	0.056	0.107	0.117	0.018	0.013	0.049	ОР100, СГ20, СГ52b

* Примечание. Алгоритмы (коды), использованные для оптимизации сетей: код ПО – псевдообратные (линейная оптимизация методом наименьших квадратов); код КС – К-средних (расстановка центров); код КБ – К-ближайших соседей (задание отклонений); код ОР – обратное распространение; код СГ – метод сопряженных градиентов; b – код остановки (сеть с наименьшей ошибкой на контрольной выборке) [44]. Код СГ31b показывает, что для оптимизации сети использован метод сопряженных градиентов и что сеть найдена на 31 эпохе по минимальной ошибке на валидационном множестве.

В качестве активационной функции на промежуточных слоях при нейросетевом моделировании использована функция гиперболического тангенса, сигмоидальная нелинейность которой определяется следующим образом [8]

$$\tanh(u) = \frac{\exp(au) - \exp(-au)}{\exp(au) + \exp(-au)},$$

где $a > 0$; $u_j = \sum_{i=1}^N w(j,i)x(i) + b(j)$ – текущее состояние нейрона (индуцированное локальное

поле нейрона j); $w(j,i)$ – вес синаптической связи выхода нейрона i с входом нейрона j , положительные значения которого соответствуют *возбуждающим* синапсам, отрицательные значения – *тормозящим* синапсам, если $w(j,i) = 0$, то связь между нейронами j и i отсутствует; $b(j)$ – порог возбуждения (активации) j -го нейрона; $x(i)$ – входные сигналы, $i = 1, 2, \dots, N$; индекс j относится к номеру рассматриваемого нейрона в сети, индекс i указывает номер синаптической связи.

Результаты оптимизации обучения многослойного персептрона МП 4:4-5-1:1 с помощью алгоритмов быстрого распространения и Левенберга-Маркара приведены в табл. 10. Эта сеть имеет лучшие статистические показатели, чем приведенная в табл. 9.

Таблица 10. Результаты * обучения персептрона МП 4:4-5-1:1

Архитектура, коэффициент корреляции Пирсона	Производительность обучения	Контрольная производительность	Тестовая производительность	Ошибка обучения	Контрольная ошибка	Тестовая ошибка	Обучение, элементы
МП 4:4-5-1:1 0.998	0.059	0.040	0.084	0.017	0.011	0.017	БР100, ЛМ235b

*Примечание: *Производительность обучения, Контрольная производительность, Тестовая производительность* – это отношение стандартного отклонения ошибки прогноза к стандартному отклонению исходных данных на соответствующих выборках (табл. 11). *Ошибка обучения, Контрольная ошибка, Тестовая ошибка* – ошибки сети на выборках, используемых во время обучения. Ошибка для конкретной конфигурации сети (мера эффективности обучения) определяется путем прогона через сеть всех имеющихся наблюдений и сравнением реально выдаваемых выходных значений (y_j) с желаемыми (целевыми) значениями (d_j). Сигнал ошибки $e_j(n)$ выходного нейрона j на итерации n , соответствующей n -му примеру обучения, рассчитывается по формуле: $e_j(n) = d_j(n) - y_j(n)$. В пакетном режиме обучения по методу обратного распространения среднеквадратическая ошибка сети определяется по уравнению [8]

$$E_{av}(n) = \frac{1}{2N} \sum_{n=1}^N \sum_{j \in C} e_j^2(n) \text{ (множество } C \text{ включает все нейроны выходного слоя сети, } N \text{ – общее число}$$

примеров в обучающем множестве). *Обучение* – на первом этапе использован алгоритм быстрого распространения (100 эпох). На втором этапе – метод Левенберга-Маркара. Сеть выбрана на 235 эпохе по минимальной ошибке на контрольной выборке (код остановки b).

Критерии качества (статистические характеристики) обученной нейросетевой модели представлены в табл. 11.

Данные табл. 12 отражают качество работы нейронной сети, в обучение которой были включены энергии Гиббса диссоциации уксусной кислоты в водно-пропан-2-ольных растворителях, в то время как результаты табл. 13 являются прогнозом силы кислоты уже обученной модели по свойствам водно-диметилсульфоксидных растворителей.

На основании анализа результатов табл. 6 и 13 можно сделать вывод о согласованности методов статистического и нейросетевого анализа термодинамики равновесий диссоциации уксусной кислоты в разных растворителях.

Таблица 11. Статистические показатели обученной нейронной сети для прогнозирования силы уксусной кислоты по свойствам водно-органических растворителей*

	Обучающая выборка $\Delta G_{d,HA}^{\circ}$	Контрольная выборка $\Delta G_{d,HA}^{\circ}$	Тестовая выборка $\Delta G_{d,HA}^{\circ}$	Общая выборка $\Delta G_{d,HA}^{\circ}$
Среднее данных	38.29	40.86	33.30	37.71
Стандартное отклонение данных	8.69	7.94	4.72	8.18
Среднее ошибки	0.009	0.053	-0.301	-0.056
Стандартное отклонение ошибки	0.51	0.31	0.40	0.47
Среднее абсолютной ошибки	0.36	0.27	0.44	0.36
Отношение стандартных отклонений	0.059	0.040	0.084	0.057
Корреляция	0.998	0.999	0.999	0.998

*Примечание. *Среднее данных* – среднее значение исходных энергий Гиббса диссоциации уксусной кислоты $\Delta G_{d,HA}^{\circ}$. В нашем случае они лежат в интервале 33 – 40, что свидетельствует о корректном разбиении массива данных на обучающую, контрольную и тестовую выборки. *Стандартное отклонение*

данных – среднееквадратическое отклонение исходных данных $\Delta G^{\circ}_{д,НАС}$. *Среднее ошибки* – среднее значение ошибки прогноза (ошибка – это разница между исходными и рассчитанными $\Delta G^{\circ}_{д,НАС}$). *Стандартное отклонение ошибки* – стандартное отклонение ошибки прогноза $\Delta G^{\circ}_{д,НАС}$. *Среднее абсолютной ошибки* – средняя абсолютная ошибка прогноза (абсолютная ошибка – это разница (по модулю) между исходными и рассчитанными $\Delta G^{\circ}_{д,НАС}$). *Отношение стандартных отклонений* или производительность сети – отношение стандартного отклонения ошибки прогноза к стандартному отклонению исходных данных (Стандартное отклонение ошибки/Стандартное отклонение данных). *Корреляция* – коэффициент корреляции Пирсона.

Таблица 12. Иллюстрация качества работы нейронной сети на примере предсказания энергии Гиббса (кДж/моль) диссоциации уксусной кислоты по свойствам водно-пропан-2-ольных растворителей

мол. доля пропан-2-ола	Свойства растворителя				$\Delta G^{\circ}_{д,НАС}$ наблюдаемые	$\Delta G^{\circ}_{д,НАС}$ предсказанные	Остатки
	$1/\varepsilon^N$	E_T^N	$B_{КТ}$	δ^2_N			
0	1.000	1.00	0.19	1.000	27.15	27.36	0.21
0.1	1.287	0.77	0.49	0.943	29.62	29.60	-0.02
0.2	1.650	0.70	0.60	0.889	32.40	32.54	0.14
0.3	2.087	0.67	0.65	0.839	35.02	34.80	-0.22
0.4	2.568	0.66	0.69	0.787	37.04	36.59	-0.45
0.5	3.032	0.64	0.72	0.730	38.49	38.21	-0.28
0.6	3.411	0.62	0.75	0.664	40.10	39.96	-0.14
0.7	3.668	0.60	0.79	0.586	42.86	43.02	0.16
0.8	3.829	0.58	0.83	0.492	47.30	46.88	-0.42
0.9	3.973	0.56	0.89	0.379	52.23	52.59	0.36
1.0	4.117	0.54	0.88	0.242	53.02	52.93	-0.09

Таблица 13. Иллюстрация качества работы нейронной сети на примере прогнозирования энергии Гиббса диссоциации уксусной кислоты по свойствам водно-диметилсульфоксидных растворителей

мол. доля DMSO	Свойства растворителя				$\Delta G^{\circ}_{д,НАС}$ наблюдаемые	$\Delta G^{\circ}_{д,НАС}$ предсказанные	Остатки
	$1/\varepsilon^N$	E_T^N	$B_{КТ}$	δ^2_N			
0	1.000	1.00	0.19	1.000	27.15	27.36	0.21
0.1	1.025	0.85	0.35	0.931	29.51	30.40	0.89
0.2	1.047	0.75	0.45	0.862	33.03	32.56	-0.47
0.3	1.091	0.69	0.50	0.793	38.83	35.13	-3.70

Однако следует отметить, что при прогнозировании констант диссоциации кислоты в водно-диметилсульфоксидных растворителях с большим содержанием диметилсульфоксида и в водно-диоксановых растворителях не удалось получить удовлетворительных результатов с помощью построенной в работе нейросети, что вполне объяснимо, так как сеть была обучена только на энергиях Гиббса диссоциации уксусной кислоты в водно-спиртовых растворителях. Одним из путей решения этой проблемы может быть пополнение массива данных по константам диссоциации кислоты и свойствам смешанных растворителей, содержащих как протолитический так и апротонный органический компонент, для обучения сети на более репрезентативных выборках и расширения ее прогностических возможностей. Для подтверждения этого заключения в работе проведен нейросетевой анализ термодинамики образования коронатов калия в растворителях вода-метанол, вода-пропан-2-ол, вода-ацетонитрил, вода-диоксан.

Изменение стандартной энергии Гиббса устойчивости коронатов калия $\Delta G^{\circ}_{LK^+}$. Анализ построенных ИНС показал, что лучшими характеристиками по производительности, точности аппроксимации, способности к обобщению (прогнозированию новых данных) и минимальной ошибке на контрольной выборке обладают нейросетевые модели персептронного типа (табл. 14).

В дальнейшем для прогнозирования устойчивости (изменения свободной энергии Гиббса, констант комплексообразования) монокоронатов калия в водно-органических растворителях использован обученный персептрон МП 4:4-5-1:1 (табл. 14) с наименее сложной архитектурой сети (сеть содержит 4 входных нейрона, пять скрытых и один выходной нейрон).

Таблица 14. Типы и характеристики нейросетевых моделей

	Архитектура, коэффициент корреляции Пирсона	Производительность обучения	Контрольная производительность	Тестовая производительность	Ошибка обучения	Контрольная ошибка	Тестовая ошибка	Обучение, элементы
1	Линейная 2:2-1:1 0.861	0.563	0.446	0.316	0.166	0.113	0.071	ПО
2	РБФ 4:4-7-1:1 0.9809	0.055	0.219	0.369	0.008	0.024	0.045	КС, КБ, ПО
3	РБФ 4:4-8-1:1 0.988	0.094	0.144	0.272	0.015	0.025	0.036	КС, КБ, ПО
4	МП 4:4-5-1:1 0.997	0.058	0.057	0.106	0.015	0.022	0.045	ОР100, СГ20, СГ66b
5	МП 4:4-8-1:1 0.997	0.048	0.038	0.104	0.012	0.022	0.040	ОР100, СГ20, СГ87b
6	МП 4:4-9-1:1 0.997	0.044	0.093	0.121	0.012	0.028	0.028	ОР100, СГ20, СГ63b

В табл. 15 и 16 приведены результаты прогноза устойчивости коронатов калия для тестовой выборки и для водно-ацетоновых растворителей, данные об энергиях Гиббса комплексообразования в которых не использовались при обучении нейросети. Константы устойчивости предсказаны с точностью $\pm (0.01 - 0.3)$ логарифмических единицы в зависимости от состава водно-органического растворителя.

Таблица 15. Предсказанные многослойным персептроном МП 4:4-5-1:1 изменения энергии Гиббса комплексообразования $\Delta G_{LK}^{\circ+}$ (МП) и остатки тестовой выборки

№*	Выборка	$1/\epsilon^N$	E_T^N	B_{KT}	δ_N^2	$\Delta G_{LK}^{\circ+}$	$\Delta G_{LK}^{\circ+}$ (МП)	Остатки
2	Тестовая	1.094	0.92	0.26	0.94	-12.99	-12.83	0.16
6	Тестовая	1.546	0.81	0.53	0.69	-21.33	-21.57	-0.24
7	Тестовая	1.679	0.80	0.54	0.63	-23.78	-23.74	0.04
10	Тестовая	2.140	0.77	0.60	0.44	-31.20	-31.26	-0.06
12	Тестовая	1.000	1.00	0.19	1.00	-11.65	-11.72	-0.07
13	Тестовая	1.287	0.77	0.49	0.94	-13.68	-14.71	-1.03
30	Тестовая	1.892	0.73	0.44	0.48	-23.32	-24.90	-1.58
31	Тестовая	1.996	0.70	0.43	0.41	-25.15	-26.68	-1.53
46	Тестовая	1.645	0.77	0.36	0.92	-15.25	-15.13	0.12

* номер независимой переменной.

В известной нам научной литературе отсутствуют экспериментальные данные по константам (энергиям Гиббса) устойчивости коронатов калия ($18\text{C}6\text{K}^+$) в водно-этанольных растворителях, но в наличии данные по физико-химическим свойствам растворителей, что позволило получить прогнозные значения энергии Гиббса (констант) устойчивости коронатов калия двумя методами – множественной линейной регрессией (МЛР) и нейросетевым оцениванием (МП) и сопоставить их (табл. 17).

Приведенные в табл. 17 результаты позволяют заключить, что оба метода в пределах изменения логарифма константы комплексообразования $\lg K \pm 0.4$ дают одинаковые прогнозные величины энергии Гиббса комплексообразования. Однако применение многослойного персептрона вместо аппарата множественной линейной регрессии позволяет существенно расширить (табл. 17) прогнозирующую способность количественных моделей, в частности, сделать прогноз и оценку его качества не зависящими от разбивки массива данных на две выборки – до области пересольватации и после области пересольватации, т.е. успешно решать задачи нелинейной аппроксимации данных.

Таблица 16. Предсказанные перцептроном изменения энергии Гиббса устойчивости коронатов калия $\Delta G_{LK}^{\circ+}$ (МП) по свойствам водно-ацетоновых растворителей в сравнении с экспериментальными данными

мол. доля А	$1/\varepsilon^N$	E_T^N	$B_{КТ}$	δ_N^2	$\Delta G_{LK}^{\circ+}$ [34]	$\Delta G_{LK}^{\circ+}$ (МП)	Остатки
0	1.000	1.00	0.19	1.000	-11.65	-11.74	-0.09
0.1	1.237	0.82	0.40	0.918	-14.43	-14.53	-0.10
0.2	1.514	0.73	0.49	0.836	-17.06	-16.66	0.40
0.3	1.820	0.70	0.52	0.753	-19.48	-19.00	0.48
0.4	2.143	0.70	0.52	0.671	-21.67	-21.88	-0.21
0.5	2.473	0.68	0.53	0.589	-23.60	-24.95	-1.35
0.55	2.637	0.66	0.56	0.548	-24.47	-26.12	-1.65

Таблица 17. Изменения энергии Гиббса устойчивости коронатов калия, предсказанные перцептроном $\Delta G_{LK}^{\circ+}$ (МП) и уравнением (2) множественной линейной регрессии (МЛР) по свойствам водно-этанольных растворителей

мол. доля EtOH	$1/\varepsilon^N$	E_T^N	$B_{КТ}$	δ_N^2	$\Delta G_{LK}^{\circ+}$ (МП)	$\Delta G_{LK}^{\circ+}$ (МЛР)	$\delta \Delta G_{LK}^{\circ+}$
0	1.000	1.00	0.19	1.000	-11.72	-11.92	0.20
0.1	1.191	0.87	0.31	0.929	-13.57	-13.83	0.26
0.2	1.405	0.78	0.51	0.859	-16.67	-15.74	-0.93
0.3	1.640	0.74	0.57	0.789	-18.86	-17.69	-1.17
0.4	1.887	0.73	0.60	0.718	-21.47	-19.64	-1.83
0.5	2.135	0.71	0.63	0.648	-23.78	-21.59	-2.19
0.6	2.371	0.70	0.66	0.578	-25.94		
0.7	2.592	0.69	0.68	0.507	-27.55		
0.8	2.799	0.67	0.71	0.437	-27.98		
0.9	3.012	0.66	0.76	0.366	-27.43		
1.0	3.224	0.66	0.77	0.296	-27.63		

ЗАКЛЮЧЕНИЕ

Актуальной задачей физикохимии растворов является создание методологии, позволяющей с единых позиций осуществлять количественный анализ и прогноз термодинамических характеристик химических равновесий на основе обработки экспериментальных данных. Показано, что такой методологией является сочетание возможностей разведочного анализа данных, линейной множественной регрессии, сольватационно-термодинамического подхода и нейросетевых технологий.

Впервые всесторонний разведочный анализ (факторный, кластерный, дискриминантный, канонический) применен к термодинамике равновесий диссоциации уксусной кислоты и комплексообразования катионов калия с эфиром 18-краун-6 в водно-органических растворителях. Построены решающие правила классификации и прогнозирования силы кислоты и устойчивости коронатов.

Установлены двухпараметрические зависимости энергии Гиббса (констант) диссоциации уксусной кислоты и устойчивости коронатов калия от диэлектрической проницаемости и плотности энергии когезии водно-органических растворителей, согласующиеся с результатами сольватационно-термодинамического анализа.

Впервые аппарат искусственных нейронных сетей применен для количественного прогнозирования силы уксусной кислоты и устойчивости коронатов калия в разных растворителях. Построены нейросетевые модели зависимости силы кислоты и устойчивости коронатов от физико-химических свойств водно-органических растворителей, обладающие высоким прогностическим потенциалом. Показано, что применение искусственных нейронных сетей вместо аппарата множественной линейной регрессии расширяет прогнозирующую способность количественных моделей.

Авторы выражают благодарность рецензентам за внимательное прочтение рукописей, конструктивные замечания и рекомендации.

ЛИТЕРАТУРА

1. Зенкин А.А. Когнитивная компьютерная графика. – М.: Наука, 1991. – 192 с.
2. Brown F.K. Chapter 35. Chemoinformatics: What is it and How does it Impact Drug Discovery // Annual Reports in Medicinal Chemistry. – 1998. – Vol. 33. – P. 375–384.
3. Leach A.R., Gillet V.J. An Introduction to Chemoinformatics. – Springer, 2007. – 256 p.
4. Bunin V.A., Siesel A., Morales G.A., Bajorath J. Chemoinformatics: Theory, Practice, & Products. – Springer, 2007. – 295 p.
5. Киреева Н.В. Прогнозирование констант устойчивости комплексов лантаноидов и щелочноземельных металлов с органическими лигандами и температур плавления ионных жидкостей методами химической информатики // Автореф. дис. ... канд. хим. наук. – М.: ИФХЭ им. А.Н. Фрумкина РАН и ун-т Страсбурга (Франция). 2010. – 25 с.
6. Пантелеймонов А.В. Новые хемометрические способы обработки данных аналитического эксперимента // Автореф. дис. ... канд. хим. наук. – Х.: – 2009 – 18 с.
7. Коняев Д.С. Методы анализа данных и химической информатики в исследовании комплексообразования в растворах и на поверхности химически модифицированных кремнезёмов. Дисс. ... канд. хим. наук. – Х.: – 1999. – с. 139–141.
8. Хайкин С. Нейронные сети: полный курс. 2-е изд.: Пер. с англ. – М.: Издательский дом "Вильямс", 2006. – 1104 с.
9. Бондарев Н.В. Сольватационно-термодинамические эффекты водно-метанольного растворителя в координации катионов Na^+ , K^+ , NH_4^+ и Ag^+ с 18-краун-6. Энергии Гиббса комплексообразования и пересольватации реагентов // Журн. общей химии. – 2006. – Т. 76, Вып. 1. – С. 13–18.
10. Смирнова Е.В., Цыба Ю.В., Бондарев Н.В., Зайцева И.С. Регрессионный анализ влияния свойств водно-диметилсульфоксидных растворителей на силу уксусной и бензойной кислот // Materiały V Międzynarodowej naukowe-praktycznej konferencji "Naukowa przestrzeń Europy - 2009". Przemysł, Polska: Nauka i studia. – 2009. – Vol. 7 (Chemia i chemiczne technologie). – S. 6–8.
11. Баскин И.И. Моделирование свойств химических соединений с использованием искусственных нейронных сетей и фрагментных дескрипторов // Автореф. дис.... докт. физ.-мат. наук. М.: МГУ им. М.В. Ломоносова. 2010. – 49 с.
12. Краснянчин Я.Н., Пантелеймонов А.В., Холин Ю.В. Надежность идентификации аналитов с помощью искусственных нейронных сетей // Вестн. Харьк. нац. ун-та. – 2010. № 895. Химия. – Вып. 18(41). – С. 39–45.
13. Баскин И.И., Палюлин В.А., Зефилов Н.С. Применение искусственных нейронных сетей в химических и биохимических исследованиях // Вестн. Моск. ун-та. Сер. 2. Химия. – 1999. – Т. 40, № 5. – С. 323–326.
14. Гальберштам Н.М., Баскин И.И., Палюлин В.А., Зефилов Н.С. Нейронные сети как метод поиска зависимостей структура - свойство органических соединений // Успехи химии. – 2003. – Т. 72, № 7. – С. 706–727.
15. Малхорта Н.К. Маркетинговые исследования. Практическое руководство, 3-е издание.: Пер. с англ. – М.: Издательский дом "Вильямс", 2002. – 960 с.
16. Наследов А.Д. SPSS: Компьютерный анализ данных в психологии и социальных науках. – СПб.: Питер, 2005. – 416 с.
17. Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. Факторный, дискриминантный и кластерный анализ: Пер. с англ. – М.: Финансы и статистика. 1989. – 216 с.
18. Боровиков В. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. 2-е изд. – СПб.: Питер, 2003. – 686 с.
19. Барсемян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсемян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. – 2-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2007. – 384 с.
20. Боровиков В.П. Нейронные сети. *Statistica Neural Networks*. Методология и технологии современного анализа данных. 2-е изд. – М.: Горячая линия – Телеком, 2008. – 392 с.

21. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного // Докл. АН СССР. – 1957. – Т. 114, № 5. – С. 953–956.
22. Tsurko E.N., Bondarev N.V. Mathematical modeling of solvent parameters' relevant contribution to the alpha-amino acid (valine, alpha-alanine) solvation in H₂O – MeOH, H₂O – EtOH and H₂O – PrOH-2 // J. Mol. Liquids. – 2007. № 131–132. – P.151–157.
23. Никольский Б.П. Справочник химика. Том 3, Изд. 2. – Л.: Изд-во «Химия», 1965. – 1008 с.
24. Лебедь В.И., Бондарев Н.В. Константы диссоциации и термодинамические характеристики диссоциации уксусной и бензойной кислот в смесях вода – метанол, вода – диоксан // Журн. физ. химии. – 1982. – Т. 56, № 1. – С. 30–33.
25. Лебедь В.И., Бондарев Н.В., Пауленова А. Константы диссоциации и термодинамические характеристики диссоциации и сольватации уксусной кислоты в смесях вода – пропанол-2 // Журн. физ. химии. – 1987. – Т. 61, № 6. – С. 1487–1491.
26. Бондарев Н.В. Количественная оценка роли водно-органического растворителя при образовании монокраун-эфирных комплексов ионов натрия и калия // Журн. физ. химии. – 1999. – Т.73, №6. – С.1019–1024.
27. Переселко В.Ф., Шевченко И.А., Жолновач А.М., Бондарев Н.В. Кондуктометрическое исследование комплексообразования иона калия с 18-краун-6 в смесях вода-метанол // Журн. общей химии. – 1995. – Т. 65, Вып. 3. – С. 363–365.
28. Кабакова Е.Н., Шевченко И.А., Жолновач А.М., Бондарев Н.В. Термодинамика комплексообразования ионов металлов с краун-эфирами. Система K⁺-18-краун-6 – вода-метанол // Журн. общей химии. – 1996. – Т. 66, Вып. 2. – С. 208–211.
29. Кабакова Е.Н., Переверзев А.Ю., Бондарев Н.В. Термодинамические параметры комплексов 18-краун-6 эфира с ионом K⁺ в смесях вода-метилловый спирт (20 и 40 % мас.) // Укр. хим. журнал. – 1996. – Т. 62, № 1. – С. 21–25.
30. Кабакова Е.Н., Бондарев Н.В. Термодинамические характеристики комплексообразования ионов Na⁺ и K⁺ с 18-краун-6 эфиром в смесях вода-метанол // Журн. физ. химии. – 1998. – Т.72, №7. – С. 1196–1199.
31. Диди Ю., Бондарев Н.В. Константы устойчивости комплекса K18K6⁺ в смесях вода-2-пропанол // Журн. общей химии. – 1996. – Т. 66, Вып. 8. – С. 1267–1270.
32. Ельцов С.В., Керн А.П., Жолновач А.М., Бондарев Н.В. Влияние растворителя вода-ацетонитрил на энергетику комплексообразования ионов Na⁺ и K⁺ с 18-краун-6 и дибензо-18-краун-6 эфирами // Журн. общей химии. – 1997. – Т. 67, Вып. 9. – С. 1430–1438.
33. Ельцов С.В., Кабакова Е.Н., Бондарев Н.В. Влияние состава смешанного растворителя вода-ацетонитрил, вода-метанол на устойчивость комплексов ионов Na⁺ и K⁺ с 18С6 эфиром // Укр. хим. журнал. – 1998. – Т.64, №4. – С.84–89.
34. Кабакова Е.Н., Бондарев Н.В. Изучение комплексообразования ионов K⁺ с 18-краун-6 в смесях вода-ацетон потенциометрическим методом при 283.15-318.15 К // Журн. неорг. химии. – 1998. – Т. 43, № 5. – С. 820–822.
35. Зайцева И.С., Ельцов С.В., Кабакова Е.Н., Бондарев Н.В. Корреляционный анализ влияния эффектов среды на энергетику комплексообразования катионов натрия и калия с эфиром 18-краун-6 в водно-органических растворителях // Журн. общей химии. – 2003. – Т.73, Вып. 7. – С. 1079–1084.
36. Райхардт К. Растворители и эффекты среды в органической химии. – М.: Мир, 1991. – 763 с.
37. Афанасьев В.Н., Ефремова Л.С., Волкова Т.В. Физико-химические свойства бинарных растворителей. Водосодержащие системы. – Иваново: ИХР РАН, 1988. – 412 с.
38. Kalidas S., Hefter G., Marcus Y. Gibbs Energies of Transfer of Cations from Water to Mixed Aqueous Organic Solvents // Chem. Rev. – 2000. – Vol. 100, № 3. – P. 819–852.
39. Hefter G., Marcus Y., Waghorne W.E. Enthalpies of Transfer of Electrolytes and Ions between Water and Mixed Aqueous Solvents // Chem. Rev. – 2002. – Vol. 102, №8. – P. 2773-2836.

40. Халафян А.А. Статистический анализ данных. STATISTICA 6.0. 2-е изд. испр. и доп. – Краснодар: КубГУ, 2005. – 308 с.
41. Гордон Д. Органическая химия растворов электролитов. – М.: Мир, 1979. – 712 с.
42. Цыба Ю.В., Бондарев Н.В., Зайцева И.С. Регрессионно-корреляционный и термодинамический анализ влияния свойств растворителей вода-метанол, вода-этанол, вода-пропан-2-ол на силу уксусной кислоты // Материали за V Международна научна практична конференция "Динамиката на съвременната наука – 2009". София, България: "Бял ГРАД-БГ" ООД. – 2009. – Том 12. (Химия и химически технологии). – С. 35–37.
43. <http://books.tr200.ru/> Иллюстрированный SPSS-самоучитель v.10.0
44. StatSoft, Inc. Электронный учебник по статистике. Москва, StatSoft. 2001. WEB: <http://www.statsoft.ru/home/textbook/default.htm>.

Поступила в редакцию 10 августа 2010 г.

С. М. Бондарев, М. В. Бондарев. Методи розвідувального аналізу і хемоінформатики в термодинаміці рівноваг.

Вперше застосовані розвідувальні методи – факторний аналіз, кластерний аналіз, дискримінантний аналіз, канонічний аналіз, побудова дерев класифікації та методи хімічної інформатики – множинна лінійна регресія, апарат нейронних мереж перцептронного типу для класифікації та прогнозування сили оцтової кислоти і стійкості коронатів калію на основі фізико-хімічних властивостей водно-органічних розчинників. Побудовано вирішальні правила класифікації і тришарові перцептрони, що дозволяють прогнозувати константи хімічних рівноваг дисоціації і комплексоутворення. Показана перспективність методології поєднання нестатистичних методів, сольватаційно-термодинамічного підходу та нейромережових технологій для виявлення закономірностей зв'язку термодинамічних характеристик хімічних рівноваг і сольватації реагентів, класифікації та прогнозування сили карбонових кислот і стійкості коронатів у різних розчинниках.

Ключові слова: розвідувальні методи, хемоінформатика, штучна нейронна мережа, перцептрон, множинна регресія, масив даних, константа рівноваги, енергія Гіббса, сольватація.

S. N. Bondarev, N. V. Bondarev. Reconnaissance Analysis and Chemoinformatics Methods in the Thermodynamics of Equilibria.

For the first time reconnaissance methods – such as factor, cluster, discriminant, and canonical analyses, classification trees construction, as well as chemical information science methods – multiple linear regression, perceptron type neural networks have been used for acetic acid dissociation constants and potassium coronates stability classification and prediction on the basis of physico-chemical properties of water-organic solvents. Determining rules and three-layer perceptrons have been constructed to predict chemical equilibrium constants of dissociation and complex formation. The potential of combining non-statistical methods with the solvation thermodynamics approach and neural network technologies for identifying regularities in the ties existing between thermodynamic characteristics of chemical equilibria and reagents solvation, classification and prediction of carboxylic acids strength and coronates stability in different solvents is explicated.

Key words: reconnaissance methods, chemoinformatics, artificial neural networks, perceptron, multiple regression, data array, equilibrium constant, Gibbs energy, solvation.

Kharkov University Bulletin. 2010. № 932. Chemical Series. Issue 19(42).