

ЭМПИРИЧЕСКИЕ МОДЕЛИ УСТОЙЧИВОСТИ КОМПЛЕКСОВ КРАУН-ЭФИРОВ С КАТИОНАМИ ЩЕЛОЧНЫХ И ЩЕЛОЧНОЗЕМЕЛЬНЫХ МЕТАЛЛОВ В НЕКОТОРЫХ ЧИСТЫХ РАСТВОРИТЕЛЯХ

© 2020 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 4 февраля 2020 г.
После доработки 4 февраля 2020 г.
Принято к печати 27 февраля 2020 г.

На основе разведочных и нейросетевых методов математического моделирования равновесий в растворах развит подход по прогнозированию констант устойчивости коронатов по свойствам растворителей, катионов и краун-эфиров. Разработаны разведочные (факторная, кластерные, дискриминантная, каноническая, деревья решения), регрессионные и нейросетевые (с учителем, сеть Кохонена) модели устойчивости комплексов краун-эфиров (12C4, 13C4, 14C4, 15C4, 15C5, 18C6, 21C7, 24C8, B12C4, B15C5, CH15C5, CH18C6, DCH18C6, DCH21C7, DB18C6, DB21C7, DB24C8, DB27C9, DB30C10) с катионами щелочных (Li^+ , Na^+ , K^+ , Cs^+ , Rb^+) и щелочноземельных (Ca^{2+} , Sr^{2+} , Ba^{2+}) металлов в водных и неводных (ацетон, ацетонитрил, диметилсульфоксид, метанол, пиридин, диметилформамид, диоксан, пропиленкарбонат, 1,2-дихлорэтан, нитробензол) растворах по свойствам растворителей (диаметр молекулы растворителя, параметр Камлета–Тафта, параметр Димрота–Райхардта, диэлектрическая проницаемость), краун-эфиров (топологический индекс Балабана) и катионов (диаметр катиона) при 298.15 К.

Ключевые слова: краун-эфиры, константа комплексообразования, разведочный анализ, множественная линейная регрессия, нейронные сети, моделирование, прогнозирование

DOI: 10.31857/S0044460X20060170

Количественные подходы к описанию влияния свойств растворителя на взаимодействие между растворенным веществом и растворителем детально рассмотрены в работах [1, 2]. В недавних работах, посвященных этой тематике, предлагается нейросетевая модель прогнозирования оптимальных условий (катализатор, растворитель, реагент, температура) для проведения конкретной органической реакции [3], разработан [4] интерактивный инструмент для выбора растворителя с использованием программного обеспечения R. Выполнено моделирование растворимости *o*-нитрофенилацетонитрила [5], D-аспарагиновой кислоты [6] в чистых растворителях, заряженных и незаряженных частиц для различных пар растворителей [7], антибиотика биापенема в разных чистых и смешанных растворителях, 2-амино-6-хлорпурина в

20 растворителях [9], *n*-нитрофенилацетонитрила в 16 растворителях при температурах от 278.15 до 333.15 К. Для 300 растворителей предложены новые эмпирические параметры, описывающие взаимодействия растворенного вещества и растворителя [11].

Методом Камлета–Тафта [1, 2] исследовано влияние водно-ацетонитрильного растворителя (50% ацетонитрила по объему) на силу (pK_{a1} и pK_{a2}) четырех противораковых препаратов – даунорубицина, доксорубицина, винкристинсульфата и 6-тиогуанина [12].

Применение пакетов прикладных программ STATISTICA и SPSS в химии растворов, клинической медицине, охране окружающей среды содействует созданию информационных технологий

Таблица 1. Описательная статистика показателей комплексообразования в разных растворителях, отобранных для разведочного анализа

Показатель	Количество значений	Среднее	Минимальное значение	Максимальное значение	Стандартное отклонение	Стандартная ошибка
$\lg K$	164	3.83	0.00	14.20	2.93	0.23
d	11	0.44	0.34	0.58	0.07	0.01
$B_{КТ}$	11	0.50	0.10	0.76	0.18	0.01
E_T	11	0.64	0.16	1.00	0.25	0.02
ϵ	11	41.00	2.21	78.36	21.14	1.65
B_{ind}	19	12.24	8.49	16.97	1.96	0.15
D_M	8	2.42	1.42	3.46	0.69	0.05

прогнозирования силы слабых электролитов [13, 14] и устойчивости краун-эфирных комплексов [15, 16] в разных растворителях, прогнозирования нарушений сердечной проводимости [17], классификации территории Украины по основным показателям повседневного функционирования [18] на основе разведочных методов анализа и нейронных сетей.

Целью данной работы являются анализ, кластеризация и разработка моделей прогнозирования термодинамических констант комплексообразования краун-эфиров с катионами металлов в зависимости от строения лиганда, взаимодействующего с ним катиона и используемого растворителя. В качестве объектов исследования выбраны комплексы краун-эфиров с катионами щелочных и щелочно-земельных металлов состава 1:1 в воде и органических средах при 298.15 К. Предметом исследования служила термодинамическая устойчивость ($\lg K$) краун-эфирных комплексов в растворах в зависимости от свойств краун-эфиров, катионов и среды, а также модели прогнозирования устойчивости коронатов. Методы исследования – математическое моделирование и методы химической информатики.

Для анализа использованы усредненные константы устойчивости коронатов катионов щелочных и щелочноземельных металлов [19–39], полученные разными авторами, в воде [19–21], метаноле [19–21, 24, 28–36], ацетонитриле [19, 24, 29, 37–39], диметилсульфоксиде [19, 24, 25, 28], ацетоне [19, 24], 1,2-дихлорэтаноле [19, 28], пиридине [19, 24, 25], *N,N*-диметилформамиде [19, 25, 28], нитробензоле [19, 28], пропиленкарбонате [19, 24], диоксане [29].

Построены корреляционные матрицы свойств растворителей, катионов и краун-эфиров. Методом

главных компонент отобраны для разработки эмпирических моделей параметры: растворителей – диэлектрическая проницаемость ϵ , параметры Димрота–Райхардта E_T и Камлета–Тафта $B_{КТ}$, диаметр молекулы растворителя d [40]; катионов – эффективный ионный кристаллохимический диаметр для координационного числа шесть D [41–43] и краун-эфиров – топологический Sum Balaban индекс [44–46]. Топологические индексы краун-эфиров рассчитаны с использованием библиотеки MathChem 1.1.3, модифицированной для работы с Python 3.5.2 [47].

Поставленная цель достигнута путем решения следующих задач: (1) первичный анализ данных, вычисление описательных статистик, проверка нормальности распределения; (2) факторный анализ – построение корреляционных матриц, выделение латентных факторов; (3) кластерный анализ – алгоритм древовидной кластеризации, итерационный алгоритм k -средних; (4) дискриминантный анализ Фишера – построение линейных классификационных функций; (5) канонический дискриминантный анализ – построение канонических линейных дискриминантных функций; (6) деревья классификации – построение правил классификации устойчивости коронатов; (7) регрессионный анализ зависимости устойчивости коронатов от свойств среды, катионов и краун-эфиров; (8) нейросетевой анализ – нейросетевой классификатор, нейросетевой аппроксиматор; (9) прогностические возможности нейросетевых моделей.

Первичный анализ данных. В табл. 1 приведены количественные параметры описательной статистики отобранных для анализа показателей. Среднее квадратическое отклонение (стандартное отклонение) данных меньше половины среднего

Таблица 2. Расчетные и табличные (критические) значения критериев проверки гипотезы нормальности распределения переменных^a

Переменная, (<i>n</i>)	Критерий Шапиро–Уилка, $W_{\text{расч}} (W_{\text{табл}})$	Критерий Колмогорова– Смирнова, $D_{\text{расч}} (D_{\text{табл}})$	Критерий Хи-квадрат Пирсона, $\chi^2_{\text{расч}} (\chi^2_{\text{табл}})$
lgK (164)		0.116 (0.069)	42.540 (16.919)
lgK* (164)		0.062 (0.069)	5.121 (9.488)
<i>d</i> (11)	0.893 (0.859)		
$B_{\text{КТ}}$ (11)	0.877 (0.859)		
E_{T} (11)	0.863 (0.859)		
ε (11)	0.798 (0.859)		
B_{ind} (19)	0.937 (901)		
<i>D</i> (8)	0.887 (0.818)		

^a *n* – объем выборки, *p* – уровень значимости. Если табличное значение $W_{\text{табл}}$ меньше расчетного значения $W_{\text{расч}}$, а $D_{\text{табл}} > D_{\text{расч}}$ и $\chi^2_{\text{табл}} > \chi^2_{\text{расч}}$, то распределение считается соответствующим нормальному на уровне значимости $p = 0.05$. Приведение данных по lgK к нормальному распределению осуществлено извлечением квадратного корня $\sqrt{\lg K} = \lg K^*$.

Таблица 3. Корреляционная матрица показателей равновесия комплексобразования

Показатели	Коэффициенты корреляции						
	lgK	<i>d</i>	$B_{\text{КТ}}$	E_{T}	ε	B_{ind}	<i>D</i>
lgK (164)	1.00	0.51	–0.69	–0.57	–0.58	0.47	0.15
<i>d</i> (11)	0.51	1.00	–0.31	–0.92	–0.62	0.18	–0.20
$B_{\text{КТ}}$ (11)	–0.69	–0.31	1.00	0.40	0.13	–0.25	–0.11
E_{T} (11)	–0.57	–0.92	0.40	1.00	0.74	–0.12	0.18
ε (11)	–0.58	–0.62	0.13	0.74	1.00	–0.20	0.12
B_{ind} (19)	0.47	0.18	–0.25	–0.12	–0.20	1.00	0.26
<i>D</i> (8)	0.15	–0.20	–0.11	0.18	0.12	0.26	1.00

арифметического, поэтому за исключением lgK и ε , распределение можно считать симметричным.

Проверка гипотезы нормального распределения анализируемых данных (табл. 2) осуществлена по критериям Шапиро–Уилка ($8 < n < 50$), Хи-квадрат Пирсона ($n > 30$) и Колмогорова–Смирнова ($n > 50$).

Факторный анализ. Надежность вычислений элементов корреляционной матрицы и целесообразность ее описания с помощью факторного анализа подтверждены мерой адекватности выборки Кайзера–Мейера–Олкина (критерий КМО = 0.562) и коэффициентом сферичности Бартлетта (критерий Хи-квадрат = 735.403, значимость критерия Бартлетта $p < 0.001$).

Методом главных компонент по выборочной совокупности значений семи отобранных показателей вычислены корреляционная матрица системы используемых для анализа данных (табл. 3), ее собственные значения, факторные нагрузки и веса факторов (табл. 4).

Свойства растворителей, краун-эфиров и катионов проявляют как умеренную положительную

(*d*, B_{ind} , *D*), так и отрицательную ($B_{\text{КТ}}$, E_{T} , ε) взаимосвязь с lgK (табл. 3). Нагрузки латентных факторов (F_1 и F_2) определены методом главных компонент с использованием критерия каменистой осыпи и процедуры ортогонального варимакс-вращения факторов. Для анализа отобрано два фактора, собственные значения которых больше единицы. Первый фактор объясняет 43.27% суммарной дисперсии, второй фактор – 26.80% (табл. 4). Переменные *d*, E_{T} и ε коррелируют с фактором 1, коэффициент корреляции равен 0.9048, –0.9504 и –0.8161 соответственно, а переменные B_{ind} и *D* коррелируют с фактором 2 (–0.7194 и –0.6700). Таким образом, первый фактор связан с вариациями свойств растворителей, а второй фактор – с изменением свойств катионов и краун-эфиров.

Математические факторные модели имеют следующий вид:

$$F_1 = 0.1251\lg K + 0.317d - 0.042B_{\text{КТ}} - 0.331E_{\text{T}} - 0.282\varepsilon - 0.061B_{\text{ind}} - 0.226D,$$

$$F_2 = -0.3231\lg K + 0.080d + 0.324B_{\text{КТ}} - 0.072E_{\text{T}} - 0.054\varepsilon - 0.406B_{\text{ind}} - 0.442D,$$

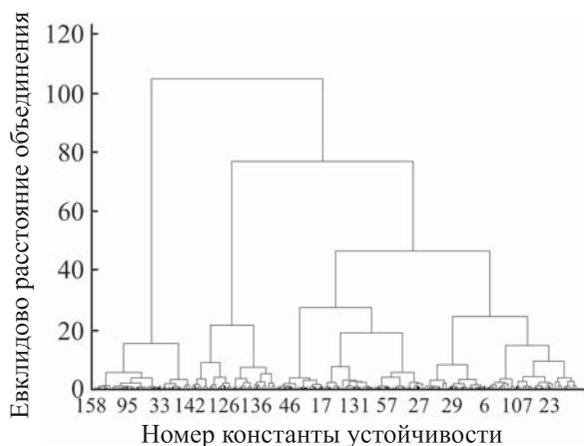
Таблица 4. Факторные нагрузки, собственные значения и веса факторов

Показатели	Факторные нагрузки	
	фактор 1 (F_1)	фактор 2 (F_2)
$\lg K$	0.6041	-0.6926
d	0.9048	-0.0720
B_{KT}	-0.3549	0.6374
E_T	-0.9504	0.0960
ε	-0.8161	0.0969
B_{ind}	0.1009	-0.7194
D	-0.3744	-0.6700
Собственные значения	3.0290	1.8762
Вес фактора, %	0.4327	0.2680

Анализ рассчитанных факторов F_1 и F_2 позволяет выяснить, какие эффекты преобладают в устойчивости коронатов в растворе – эффекты среды или свойства катионов и краун-эфиров.

Кластерный анализ. В работе реализованы два метода кластерного анализа: агломеративный – объединение, или дерево кластеризации и дивизивный – кластеризация k -средними. Предварительно была проведена процедура стандартизации исходных данных (z -оценки) путем вычитания среднего и деления на стандартное отклонение.

Агломеративная кластеризация. На рис. 1 приведена дендрограмма иерархической кластеризации устойчивости 164 коронатов по свойствам растворителей, катионов и краун-эфиров.

**Рис. 1.** Дендрограмма иерархической кластеризации констант устойчивости коронатов.

Объединение констант устойчивости коронатов в кластеры проведено методом Варда с использованием евклидова расстояния в качестве метрики пространства. На расстоянии, равном 40, выявлено 4 кластера; при увеличении расстояния до 60 количество кластеров равно трем, на расстоянии 80 – 2 кластера.

Кластерный анализ алгоритмом k -средних.

Наилучшее согласие результатов двух методов кластерного анализа получено при выборе 4 кластеров. На рис. 2 приведен график средних значений показателей комплексообразования в четырех кластерах, отображающих различие между группами констант устойчивости коронатов по каждому из свойств.

Результаты дисперсионного анализа свидетельствуют (табл. 5), что распределение констант устойчивости по кластерам проведено успешно. Уровень значимости p у критерия Фишера значительно меньше 0.05 для всех переменных, а наблюдаемый критерий Фишера больше критического $F_{набл} > F_{кр}$.

Количественный (164 константы) состав кластеров: первый кластер объединяет 27 констант устойчивости коронатов в апротонных растворителях диоксане, 1,2-дихлорэтане, нитробензоле; второй кластер – 33 константы устойчивости коронатов в воде – протолитическом растворителе; третий кластер группирует 56 констант устойчивости коронатов в апротонных (MeCN, ДМСО, пропи-

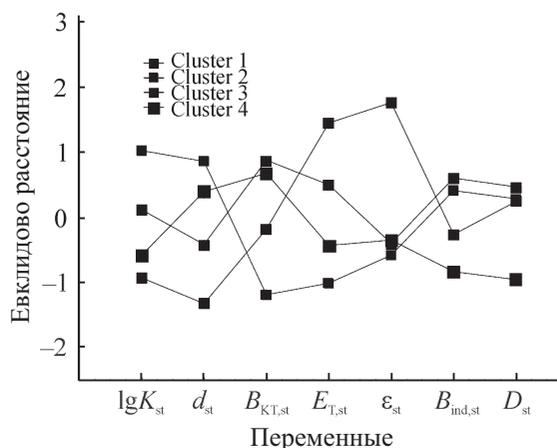
**Рис. 2.** Средние значения показателей комплексообразования для четырех групп констант устойчивости коронатов катионов щелочных и щелочноземельных металлов.

Таблица 5. Результаты дисперсионного анализа стандартизованных показателей комплексообразования методом k -средних^a

Показатель	σ_1^2	ν_1	σ_2^2	ν_2	$F(3, 160)$	p
$\lg K_{st}$	111.45	3	51.55	160	115.29	0.000
d_{st}	122.80	3	40.20	160	162.91	0.000
$B_{КТ, st}$	103.65	3	59.35	160	93.15	0.000
$E_{Т, st}$	119.52	3	43.48	160	146.61	0.000
ε_{st}	137.70	3	25.30	160	290.33	0.000
$B_{ind, st}$	57.84	3	105.16	160	29.33	0.000
D_{st}	73.49	3	89.51	160	43.78	0.000

^a σ_1^2 – межгрупповая дисперсия; σ_2^2 – внутригрупповая дисперсия; ν_1, ν_2 – степени свободы; $F(3, 160)$ – наблюдаемый критерий Фишера, [$F_{кр}(3, 160, p 0.05) = 2.66$]; p – наблюдаемый уровень значимости.

Таблица 6. Результаты дискриминантного анализа (алгоритм – переменные в модели)

Свойство	Группирующая переменная: 4 кластера констант устойчивости коронатов; Λ -Уилкса: 0.031; $F_{набл}(18, 438) = 163.36, p < 0.000; F_{кр}(18, 438) = 1.54, F_{кр}(3, 155) = 2.66$					
	Λ -Уилкса	частная Λ -Уилкса	$F_{искл}(3, 155)$	p -уровень	толерант- ность, $1-R^2$	R^2 ^a
d	0.014	0.216	187.979	0.000	0.116	0.884
$B_{КТ}$	0.019	0.164	262.529	0.000	0.295	0.705
$E_{Т}$	0.009	0.361	91.390	0.000	0.114	0.886
ε	0.017	0.186	225.977	0.000	0.423	0.577
B_{ind}	0.004	0.837	10.041	0.000	0.965	0.035
D	0.005	0.637	29.453	0.000	0.987	0.013

^a R^2 – коэффициент детерминации.

ленкарбонат) и протолитическом (MeOH) растворителях и четвертый кластер – 48 констант устойчивости коронатов в апротонных (MeCN, ДМСО, пропиленкарбонат, пиридин, ацетон, ДМФА) и протолитическом (MeOH) растворителях [48].

Распределение девятнадцати краун-эфиров по кластерам: 1, 2, 3, 4 кластеры – 18C6, DB18C6; 2, 3, 4 кластеры – 15C5, CH15C5, CH18C6, DCH18C6; 1, 3 кластеры – DB24C8, DB30C10; 3, 4 кластеры – B15C5; 2 кластер – DCH21C7; 3 кластер – 21C7, 24C8, DB21C7, DB27C9; 4 кластер – 12C4, 13C4, 14C4, 15C4, B12C4.

Распределение катионов по кластерам: 1, 2 кластеры – все участвующие в комплексообразовании восемь катионов; 3 кластер – $\text{Na}^+, \text{K}^+, \text{Rb}^+, \text{Cs}^+, \text{Ca}^{2+}, \text{Sr}^{2+}, \text{Ba}^{2+}$; 4 кластер – $\text{Li}^+, \text{Na}^+, \text{K}^+$.

Для подтверждения результатов кластерного анализа проведен дискриминантный и канонический анализ влияния свойств растворителей, катионов и краун-эфиров на устойчивость коронатов, построены деревья классификации и нейросетевые аппроксиматор и классификаторы.

Дискриминантный анализ. Для разделения констант устойчивости коронатов на группы по свойствам растворителей, краун-эфиров и катионов проведен линейный дискриминантный анализ Фишера, реализованный в статистическом пакете STATISTICA 12.

Значение стандартной статистики Уилкса лямбда (Λ -Уилкса) равно 0.031 (табл. 6), что свидетельствует о высокой дискриминирующей мощности модели (1.0 – дискриминация отсутствует, 0.0 – полная дискриминация). Этот вывод также подтверждается наблюдаемым значением $F_{набл}$ -статистики, $F_{набл}(18, 438) = 163.36, p < 0.000$ и $F_{набл}(18, 438) > F_{кр}(18, 438)$. Из табл. 6 (столбец Λ -Уилкса) также следует, что только три свойства – $B_{КТ}$, ε и d обладают наибольшей мощностью дискриминации, чем больше значение Λ -Уилкса, тем более желателен параметер в процедуре разделения констант устойчивости коронатов на группы. Частная Λ -Уилкса, характеризующая единичный вклад соответствующего свойства в дискриминирующую мощь модели, подтверждает этот вывод. Чем меньше значение частной Λ -Уилкса, тем боль-

Таблица 7. Матрица кластеризации констант устойчивости коронатов^a

Кластер	% правильной кластеризации	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Кластер 1	100.0	27	0	0	0
Кластер 2	100.0	0	33	0	0
Кластер 3	98.2	0	0	55	1
Кластер 4	97.9	0	0	1	47
Всего, %	98.8	27	33	56	48

^a Строки матрицы – наблюдаемая кластеризация методом k -средних. Столбцы матрицы – предсказанная классификация дискриминантным анализом Фишера.

Таблица 8. Характеристика извлеченных канонических корней (канонических линейных дискриминантных функций)^a

Извлеченные корни	Co	R	R^2	Λ	χ^2	ν	p
0	24.40	0.9801	0.9606	0.0031	912.19	18	0.000
1	4.79	0.9096	0.8274	0.0790	401.07	10	0.000
2	1.19	0.7365	0.5424	0.4576	123.51	4	0.000

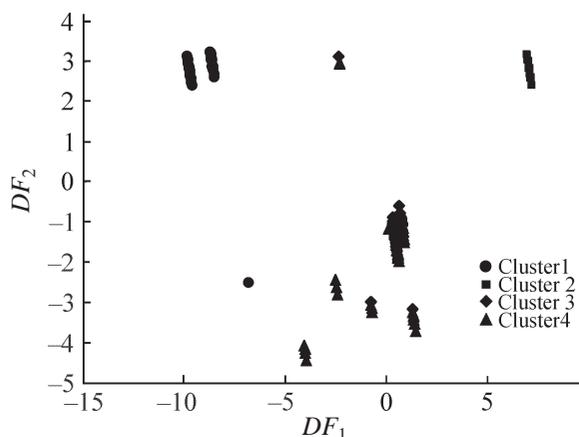
^a Co – собственное значение, R – коэффициент канонической корреляции, R^2 – коэффициент детерминации, Λ – значение статистики Λ -Уилкса, χ^2 – значение статистики Хи-квадрат, ν – число степеней свободы, p – уровень значимости соответствующего канонического корня.

ший вклад этого свойства в общую дискриминацию. Наряду с этим, чем меньше значение критерия Фишера $F_{\text{искл}}$ (табл. 6), тем менее желательны свойства в модели дискриминации – E_T , D , B_{ind} . Значения толерантности, близкие к единице, свидетельствуют об избыточности переменных B_{ind} и D в дискриминантной модели.

Математические модели дискриминации констант устойчивости коронатов – линейные классификационные функции имеют следующий вид:

$$\text{Кластер 1} = 804.222d + 2.702B_{\text{КТ}} + 363.138E_T - 0.306\epsilon + 3.513B_{\text{ind}} + 10.872D - 344.450;$$

$$\text{Кластер 2} = 819.224d + 26.656B_{\text{КТ}} + 312.827E_T - 0.840\epsilon + 2.990B_{\text{ind}} + 12.179D - 292.998;$$

**Рис. 3.** Распределение констант устойчивости коронатов по четырем кластерам в координатах DF_1 – DF_2 .

$$\text{Кластер 3} = 856.128d + 82.774B_{\text{КТ}} + 361.550E_T - 1.235\epsilon + 3.957B_{\text{ind}} + 11.533D - 363.431;$$

$$\text{Кластер 4} = 837.835d + 79.233B_{\text{КТ}} + 325.394E_T - 1.100\epsilon + 2.455B_{\text{ind}} + 8.519D - 309.325.$$

Подставив в эти уравнения значения свойств растворителя, катиона и краун-эфира, которые не использовались при построении линейных классификационных функций, можно предсказать кластер (первый, второй, третий или четвертый), к которому константа устойчивости будет отнесена по наибольшему рассчитанному значению линейной классификационной функции [15].

Высокую мощность дискриминантной модели демонстрирует матрица классификации (табл. 7). На диагонали матрицы содержится количество констант устойчивости коронатов правильно классифицированных в кластеры. В третий кластер правильно отнесены 55 констант устойчивости коронатов из 56, что составляет 98.2% правильной кластеризации. Одна константа устойчивости ошибочно отнесена к четвертому кластеру. В четвертый кластер правильно отнесены 47 констант устойчивости – 97.9% правильной кластеризации. Одна константа устойчивости ошибочно отнесена к третьему кластеру. Таким образом, дискриминантная модель кластеризации констант устойчивости коронатов на 98.8% подтвердила результаты метода k -средних кластерной модели.

Канонический анализ. Показано как канонические линейные дискриминантные функции не-

зависимых свойств растворителей, краун-эфиров и катионов распределяют константы устойчивости коронатов по 4 кластерам, выделенным методом k -средних.

Вычислены три независимые (ортогональные) дискриминирующие функции (табл. 8). Первая строка содержит критерий значимости для всех дискриминантных функций (корней), во второй строке приведены данные о значимости дискриминантных функций, оставшихся после удаления первой, в третьей строке содержатся данные о значимости функций, оставшихся после удаления первых двух. Каждая последующая дискриминантная функция вносит все меньший и меньший вклад в общую дискриминацию. Из анализа данных табл. 8 вытекает, что извлеченные три канонические корни (дискриминантные функции) статистически значимы, так как собственные значения больше единицы, а значения канонической корреляции равны 0.9801, 0.9096 и 0.7365 соответственно.

Математические канонические модели для расчета канонических линейных дискриминантных функций (DF) имеют следующий вид:

$$DF_1 = -68.241d + 10.588B_{\text{КТ}} - 14.602E_{\text{T}} + 0.166\epsilon - 0.023B_{\text{ind}} - 0.062D + 27.429$$

$$DF_2 = 28.860d - 12.412B_{\text{КТ}} + 12.085E_{\text{T}} + 0.023\epsilon + 0.086B_{\text{ind}} + 0.176D - 16.493$$

$$DF_3 = 10.852d - 1.962B_{\text{КТ}} + 1.979E_{\text{T}} + 0.018\epsilon - 0.333B_{\text{ind}} - 1.571D + 2.111$$

Константа устойчивости исследуемого короната, для которого по свойствам растворителя, краун-эфира и катиона рассчитаны канонические линейные дискриминантные функции DF_1 , DF_2 и

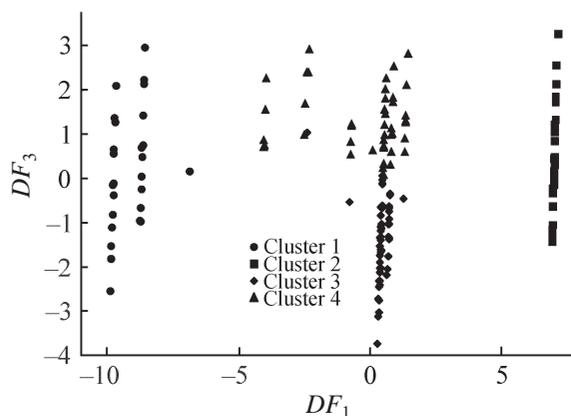


Рис. 4. Распределение констант устойчивости коронатов по четырем кластерам в координатах DF_1 – DF_3 .

Таблица 9. Средние канонических переменных (центроиды кластеров)

Кластер	DF_1	DF_2	DF_3
Кластер 1	-9.166	2.675	0.133
Кластер 2	7.015	2.904	0.339
Кластер 3	0.443	-1.174	-1.371
Кластер 4	-0.184	-2.132	1.292

DF_3 будет отнесена к кластеру по наименьшему расстоянию до центра (центроида) соответствующего кластера [15]. В табл. 9 приведены координаты центроидов четырех кластеров констант устойчивости коронатов.

Первая дискриминантная функция DF_1 отделяет первый кластер от второго (рис. 3, 4), в то время как третий и четвертый кластеры разделяются третьей дискриминантной функцией DF_3 (рис. 4). Вторая дискриминантная функция DF_2 отделяет первый и второй кластеры от третьего и четвертого, а третий и четвертый кластеры разделяются третьей дискриминантной функцией DF_3 (рис. 5).

Деревья классификации. Методологические аспекты построения деревьев классификации (правил решения) алгоритмом CART (classification and regression trees) изложены в работах [13, 15]. На рис. 6 приведен граф дерева классификации устойчивости коронатов катионов щелочных и щелочноземельных металлов – четыре вершины ветвления (1, 2, 5, 6) и пять терминальных вершин (3, 4, 7, 8, 9) – обозначения в верхнем левом углу вершин. Текст под вершинами ветвления описывает условие ветвления. Числа в правом верхнем углу вершин обозначают номер кластера. Числа

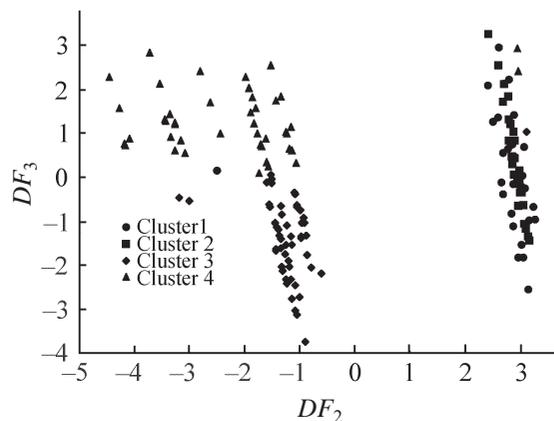


Рис. 5. Распределение констант устойчивости коронатов по четырем кластерам в координатах DF_2 – DF_3 .

Таблица 10. Структура дерева классификации устойчивости коронатов

Вершина	Левая вершина	Правая вершина	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Предсказанный кластер	Константа ветвления	Переменная ветвления
1	2	3	27	33	56	48	3	0.881	E_T
2	4	5	27	0	56	48	3	0.385	B_{KT}
3			0	33	0	0	2		
4			27	0	0	0	1		
5	6	7	0	0	56	48	3	2.0	D_M
6	8	9	0	0	5	46	4	12.865	B_{ind}
7			0	0	51	2	3		
8			0	0	1	46	4		
9			0	0	4	0	3		

над вершинами показывают количество констант устойчивости коронатов, отнесенных к данной вершине. Все константы устойчивости коронатов в вершинах ветвления относятся к кластеру, в котором количество констант устойчивости наибольшее. Поэтому корневая вершина ветвления 1 обозначена как Cluster 3.

В табл. 10 приведена структура дерева классификации устойчивости коронатов катионов по свойствам растворителей (E_T , B_{KT}), катионов (D_M) и краун-эфиров (B_{ind}). Ранги значимости предикторов дерева кластеризации d , B_{KT} , E_T , ε , D_M , B_{ind} равны 81, 81, 89, 71, 100, 54 соответственно (0 – низкая значимость, 100 – высокая значимость).

Проведенный кластер-анализ с использованием деревьев решений на 98.5% подтвердил (рис. 6, табл. 10) результаты кластерного анализа устойчивости коронатов методом k -средних. Состав первого и второго кластеров подтверждены на 100%. Одна константа третьего кластера алгоритмом CART ошибочно отнесена в четвертый кластер (98.2%), а две константы четвертого кластера ошибочно отнесены к третьему кластеру (95.8%).

Регрессионный анализ. Математические регрессионные модели [49, 50] имеют следующий вид:

– включены все переменные

$$\lg K = (1.29 \pm 4.57) + (9.89 \pm 7.90)d - (10.16 \pm 1.43)B_{KT} + (3.77 \pm 2.84)E_T - (0.08 \pm 0.02)\varepsilon + (0.26 \pm 0.12)B_{ind} + (0.42 \pm 0.31)D_M,$$

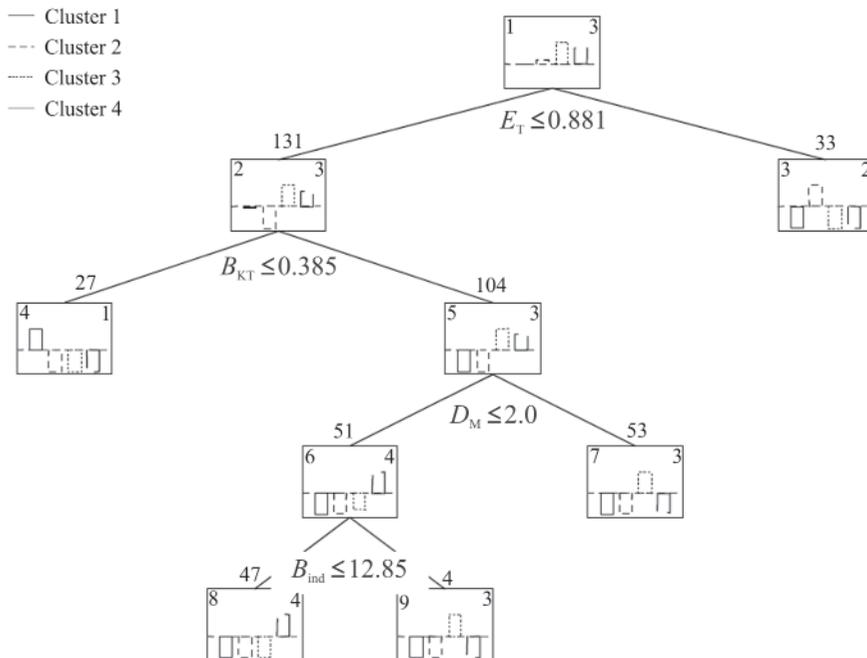


Рис. 6. Граф дерева классификации устойчивости коронатов катионов щелочных и щелочноземельных металлов.

Таблица 11. Итоги нейросетевого аппроксиматора MLP 6-8-1^a

Архитектура	Производительность обучения	Контрольная производительность	Тестовая производительность	Ошибка обучения	Контрольная ошибка	Тестовая ошибка	Алгоритм обучения	Функция ошибки	Функция активации скрытых нейронов	Функция активации выходных нейронов
MLP 6-8-1	0.962	0.965	0.965	0.334	0.200	0.284	BFGS 51	SOS	Exponent	Logistic

^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 – ошибки сети на соответствующих выборках; BFGS – Алгоритм Бройдена–Флетчера–Гольдфарба–Шанно; SOS – среднеквадратичная ошибка $E = \frac{1}{P} \sum_{i=1}^P (\lg K_{\text{расч},i} - \lg K_{\text{эсп},i})^2$, P – количество обработанных примеров в выборке; Exponent – экспоненциальная функция $\varphi(x) = e^x$; Logistic – логистическая функция $\varphi(x) = 1/[1 + \exp(-tx)]$.

Таблица 12. Основные характеристики нейросетевых классификаторов RBF 6-27-4 и MLP 6-7-4^a

Архитектура	Производительность обучения	Контрольная производительность	Тестовая производительность	Алгоритм обучения	Функция ошибки	Функция активации скрытых нейронов	Функция активации выходных нейронов
RBF 6-27-4	97.4	100	100	RBFT	SOS	Gaussian	Identical
MLP 6-7-4	98.3	100	100	BFGS 15	SOS	Logistic	Tanh

^a RBFT – обучающий алгоритм радиальной базисной функции; Gaussian – $\varphi(x) = \exp(-x^2)$; Tanh – гиперболический тангенс $\text{th}(Ax) = (e^{Ax} - e^{-Ax}) / (e^{Ax} + e^{-Ax})$; Identical – тождественная функция $\varphi(x) = x$.

$R = 0.8932$, наблюдаемое значение критерия Фишера $F_{\text{набл}}(6, 157) = 103.25$, $p < 0.000$, критическое значение критерия Фишера $F_{\text{кр}}(6, 157) = 2.16$, $p = 0.05$, стандартная ошибка = 1.34, критерий Дарбина–Уотсона $d_{\text{DW}} = 1.91$;

– отбор переменных методом прямого выбора (forward selection)

$$\lg K = (6.46 \pm 1.77) - (9.35 \pm 1.22)B_{\text{КТ}} - (0.07 \pm 0.02)\varepsilon + (0.31 \pm 0.12)B_{\text{инд}} + (0.39 \pm 0.32)D_{\text{М}},$$

$$R = 0.8882, F_{\text{набл}}(4, 159) = 148.51, p < 0.000,$$

$$F_{\text{кр}}(4, 159) = 2.43, p = 0.05,$$

стандартная ошибка = 1.36, $d_{\text{DW}} = 1.86$;

– отбор переменных методом обратного исключения (backward elimination)

$$\lg K = (6.89 \pm 1.77) - (9.45 \pm 1.24)B_{\text{КТ}} - (0.06 \pm 0.01)\varepsilon + (0.35 \pm 0.12)B_{\text{инд}},$$

$$R = 0.8838, F_{\text{набл}}(3, 160) = 190.29, p < 0.000,$$

$$F_{\text{кр}}(3, 160) = 2.66, p = 0.05,$$

стандартная ошибка = 1.39, $d_{\text{DW}} = 1.85$.

Результаты разведочных и регрессионных методов анализа данных с высокой согласованностью подтвердили статистическую значимость выбранных входных переменных для построения прогностических нейросетевых моделей зависимости констант устойчивости коронатов $\lg K$ от свойств растворителей, катионов и краун-эфиров.

Нейросетевой анализ. В табл. 11 приведены основные характеристики обученного нейросетевого аппроксиматора – многослойного перцептрона MLP 6-8-1. Коэффициенты корреляции на обучающей (70%), контрольной (15%) и тестовой (15%) выборках равны 0.9620, 0.9647 и 0.9659

Таблица 13. Основные характеристики сети Кохонена как классификатора

Архитектура	lgK устойчивости коронатов катионов							
	Ошибка			Алгоритм обучения	Количественный состав кластеров			
	Обучающая выборка 70%	Контрольная выборка 15 %	Тестовая выборка 15%		Кластер 1	Кластер 2	Кластер 3	Кластер 4
SOFM 10-4	0.2028	0.2201	0.1561	Kohonen 1000	27	33	56	48

Таблица 14. Итоги кластеризации констант устойчивости коронатов нейронными сетями: радиальной базисной функцией RBF 6-27-4 и многослойным персептроном MLP 6-7-4

Архитектура сети	Показатели кластеризации	Кластер 1	Кластер 2	Кластер 3	Кластер 4	Все
RBF 6-27-4	Все	27	33	56	48	164
	Правильно	27	33	55	46	161
	Неправильно	0	0	1	2	3
	Правильно, %	100	100	98	96	98.17
	Неправильно, %	0	0	2	4	1.83
MLP 6-7-4	Все	27	33	56	48	164
	Правильно	27	33	55	47	162
	Неправильно	0	0	1	1	2
	Правильно, %	100	100	98	98	98.78
	Неправильно, %	0	0	2	2	1.22

Таблица 15. Предсказанные персептроном MLP 6-8-1 значения lgK устойчивости коронатов

Краун-эфир	Катион	Растворитель	lgK _{эксп}	lgK _{MLP}	Краун-эфир	Катион	Растворитель	lgK _{эксп}	lgK _{MLP}
12C4	Mg ²⁺	Пропиленкарбонат	2.61	1.27	DB18C6	Ag ⁺	Вода	1.41	1.86
15C5	Ag ⁺	Вода	0.94	1.53	DB18C6	Tl ⁺	Вода	1.50	1.75
15C5	Tl ⁺	Вода	1.23	1.08	DB18C6	Pb ²⁺	Вода	1.89	1.93
15C5	Pb ²⁺	Вода	1.85	1.56	DB18C6	Ag ⁺	MeOH	4.04	3.93
18C6	Ag ⁺	Вода	1.55	1.90	DB18C6	Tl ⁺	MeCN	4.90	4.95
18C6	Tl ⁺	Вода	2.27	2.10	DB21C7	Tl ⁺	MeCN	>5.00	5.36
18C6	Pb ²⁺	Вода	4.30	2.00	DB21C7	Tl ⁺	Ацетон	4.71	4.83
18C6	Ag ⁺	MeOH	4.57	3.96	DB21C7	Tl ⁺	MeOH	3.97	4.38

соответственно. Статистические характеристики обученной нейросетевой модели персептронного типа MLP 6-8-1 (табл. 11) отражают успешность проведенного обучения. Так, качество обучения на различных выборках больше 96%, ошибка обучения на контрольной выборке (0.200) не превышает ошибку на независимой тестовой выборке (0.284). Эти данные также свидетельствуют о том, что нейросетевая модель обладает большей прогнозирующей силой, чем модели множественной линейной регрессии, коэффициенты корреляции которых не превышают 0.9.

В табл. 12 и 13 приведены основные характеристики обученных персептронных классификаторов – RBF 6-27-4, MLP 6-7-4 и SOFM 10-4. Алгоритмы радиальной базисной функции, многослойного персептрона и сети Кохонена, соответственно, на 98.17, 98.78 и 100% подтвердили правомочность кластеризации методом *k*-средних (табл. 13, 14).

Прогностические возможности персептрона. Табл. 15 демонстрирует возможности обученного многослойного персептрона MLP 6-8-1, как аппроксиматора, по предсказанию констант устойчивости коронатов lgK_{MLP} по свойствам

растворителей, катионов и краун-эфиров. Экспериментальные константы комплексообразования $\lg K_{\text{эксп}}$, взятые из работы [19], не участвовали в обучении нейронной сети.

На основе известных методов математического моделирования разработаны методология и модели количественного описания влияния свойств растворителя, катионов и краун-эфиров на устойчивость комплексов макроциклических лигандов с катионами щелочных и щелочноземельных металлов. Построены прогностические модели константа комплексообразования–свойство для кластеризации, расчета и прогнозирования констант устойчивости коронатов.

Автор выражает благодарность IT-специалисту Д.А. Козлову за помощь в обработке исходных данных по константам устойчивости коронатов и расчеты топологических индексов краун-эфиров.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов

СПИСОК ЛИТЕРАТУРЫ

1. Kamlet M.J., Abboud J.L.M., Abraham M.H., Taft R.W. // J. Org. Chem. 1983. Vol. 48. N 17. P. 2877. doi 10.1021/jo00165a018
2. Politzer P., Murray J.S. Quantitative Approaches to Solute-Solvent Interactions. Modern Aspects of Electrochemistry / Ed. C. Vayenas. New York: Springer, 2005. N 39. P. 1. doi 10.1007/978-0-387-31701-4_1
3. Gao H., Struble T.J., Coley C.W., Wang Y., Green W.H., Jensen K.F. // ACS Cent. Sci. 2018. Vol. 4. N 11. P. 1465. doi 10.1021/acscentsci.8b00357
4. Piccione P.M., Baumeister J., Salvesen T., Flores Y., Grosjean Ch., Murudi V., Shyadligeri A., Lobanova O., Lothschütz Ch. // Org. Proc. Res. Dev. 2019. Vol. 23. N 5. P. 998. doi 10.1021/acs.oprd.9b00065
5. Wang H., Wang X., Chen G., Farajtabar A., Zhao H., Li X. // J. Chem. Eng. Data. 2019. Vol. 64. N 6. P. 2867. doi 10.1021/acs.jced.9b00243
6. Wu J., Wang J., Zhao H. // J. Chem. Eng. Data. 2019. Vol. 64. N 6. P. 2904. doi 10.1021/acs.jced.9b00320
7. Qiu J., Albrecht J. // Org. Process Res. Dev. 2018. Vol. 22. N 7. P. 829. doi 10.1021/acs.oprd.8b00117
8. Xu R., Huang C., Xu J. // J. Chem. Eng. Data. 2019. Vol. 64. N 4. P. 1454. doi 10.1021/acs.jced.8b01051
9. Li W., Zhu Ya., Wang X., Zheng M., Li X., Zhao H. // J. Chem. Eng. Data. 2019. Vol. 64. N 2. P. 771. doi 10.1021/acs.jced.8b01014
10. Chen G., Liang J., Han J., Zhao H. // J. Chem. Eng. Data. 2019. Vol. 64. N 1. P. 315. doi 10.1021/acs.jced.8b00811
11. Laurence C., Legros J., Chantzis A., Planchat A., Jacquemin D. // J. Phys. Chem. B. 2015. Vol. 119. N 7. P. 3174. doi 10.1021/jp512372c
12. Sanli S., Altun Y., Guven G. // J. Chem. Eng. Data. 2014. Vol. 59. N 12. P. 4015. doi:10.1021/je500595w
13. Бондарев Н.В. // ЖОХ. 2016. Т. 86. № 6. С. 887; Bondarev N.V. // Russ. J. Gen. Chem. 2016. Vol. 86. N 6. P. 1221. doi 10.1134/S1070363216060025
14. Бондарев Н.В. // ЖОХ. 2017. Т. 87. № 2. С. 207; Bondarev N.V. // Russ. J. Gen. Chem. 2017. Vol. 87. N 2. P. 188. doi 10.1134/S1070363217020062
15. Бондарев Н.В. // ЖОХ. 2019. Т. 89. № 2. С. 288; Bondarev N.V. // Russ. J. Gen. Chem. 2019. Vol. 89. N 2. P. 281. doi 10.1134/S1070363219020191
16. Бондарев Н.В. // ЖОХ. 2019. Т. 89. № 7. С. 1085; Bondarev N.V. // Russ. J. Gen. Chem. 2019. Vol. 89. N 7. P. 1438. doi 10.1134/S1070363219070144
17. Никулина С.Ю., Черноva А.А., Третьякова С.С., Никулин Д.А. // Рос. кардиол. ж. 2018. Т. 23. № 10. С. 53. doi 10.15829/1560-4071-2018-10-53-58
18. Тютюник В.В., Бондарев М.В., Шевченко Р.И., Чорногор Л.Ф., Калугин В.Д. // Техногенно-екологічна безпека та цивільний захист. 2014. № 7. С. 107.
19. Izatt R.M., Bradshaw J.S., Nielsen S.A., Lamb J.D., Christensen J. J., Sen D. // Chem. Rev. 1985. Vol. 85. N 4. P. 271. doi 10.1021/cr00068a003
20. Frensdorff H. K. // J. Am. Chem. Soc. 1971. Vol. 93. N 3. P. 600. doi 10.1021/ja00732a007
21. Pedersen C.J., Frensdorff H.K. // Angew Chem. Int. Ed. 1972. Vol. 11. N 1. P. 16. doi 10.1002/anie.197200161
22. Izatt R.M., Terry R.E., Haymore B.L., Hansen L.D., Dalley N.K., Avondet A.G., Christensen J.J. // J. Am. Chem. Soc. 1976. Vol. 98. N 24. P. 7620. doi 10.1021/ja00440a028
23. Høiland H., Ringseth J.A., Brun T.S. // J. Solut. Chem. 1979. Vol. 8. N 11. P. 779. doi 10.1007/bf00648577
24. Smetana A.J., Popov A.I. // J. Solut. Chem. 1980. Vol. 9. N 3. P. 183. doi 10.1007/bf00648325
25. Lin J.D., Popov A.I. // J. Am. Chem. Soc. 1981. Vol. 103. N 13. P. 3773. doi 10.1021/ja00403a026
26. Host Guest Complex Chemistry III / Eds F. Vögtle, E. Weber. Berlin; Heidelberg: Springer, 1984. Vol. 121. P. 1. doi 10.1007/3-540-12821-2_1
27. Stover F.S. // J. Chromatogr. (A). 1984. Vol. 298. P. 203. doi 10.1016/s0021-9673(01)92714-1
28. Samec Z., Papoff P. // Anal. Chem. 1990. Vol. 62 N 10. P. 1010. doi 10.1021/ac00209a009
29. Gokel G.W., Leevy W.M., Weber M.E. // Chem. Rev. 2004. Vol. 104. N 5. P. 2723. doi 10.1021/cr020080k

30. *Bondarev H.B.* // ЖОХ. 2006. Т. 76. № 1. С. 13; *Bondarev N.V.* // Russ. J. Gen. Chem. 2006. Vol. 76. N 1. P. 11. doi 10.1134/S1070363219020191
31. *Gokel G.W., Goli D.M., Minganti C., Echegoyen L.* // J. Am. Chem. Soc. 1983. Vol. 105. N 23. P. 6786. doi 10.1021/ja00361a003
32. *Inoue Y., Hakushi T., Liu Y., Tong L.H.* // J. Org. Chem. 1993. Vol. 58. N 20. P. 5411. doi 10.1021/jo00072a024
33. *Lamb J.D., Izatt R.M., Swain C.S., Christensen J.J.* // J. Am. Chem. Soc. 1980. Vol. 102. N 2. P. 475. doi 10.1021/ja00522a005
34. *Haymore B.L., Lamb J.D., Izatt R.M., Christensen J.J.* // Inorg. Chem. 1982. Vol. 21. N 4. P. 1598. doi 10.1021/ic00134a065
35. *Bradshaw J.S., Izatt R.M.* // Acc. Chem. Res. 1997. Vol. 30. N 8. P. 338. doi 10.1021/ar950211m
36. *Shamsipur M., Popov A.I.* // J. Am. Chem. Soc. 1979. Vol. 101. N 15. P. 4051. doi 10.1021/ja00509a005
37. *Hopkins H.P., Norman A.B.* // J. Phys. Chem. 1980. Vol. 84. N 3. P. 309. doi 10.1021/j100440a019
38. *Buschmann H.-J.* // J. Sol. Chem. 1988. Vol. 17. N 3. P. 277. doi 10.1007/bf00646180
39. *Ельцов С.В., Дорошенко А.О., Бондарев Н.В.* // ЖНХ. 1999. Т. 44. № 2. С. 329; *El'tsov S.V., Doroshenko A.O., Bondarev N.V.* // Russ. J. Inorg. Chem. 1999. Vol. 44. N 2. P. 284.
40. *Marcus Y.* The Properties of Solvents. Chichester: John Wiley & Sons, 1999. Vol. 4. 399 p.
41. *Shannon R.D., Prewitt C.T.* // Acta Crystallogr. (B). 1969. Vol. 25. N 5. P. 925. doi 10.1107/s0567740869003220
42. *Shannon R.D., Prewitt C.T.* // J. Inorg. Nucl. Chem. 1970. Vol. 32. N 5. P. 1427. doi 10.1016/0022-1902(70)80629-7
43. *Бугаенко Л.Т., Рябых С.М., Бугаенко А.Л.* // Вестн. МГУ. Сер. Химия. 2008. Т. 49. № 6. С. 363; *Bugaenko L.T., Ryabykh S.M., Bugaenko A.L.* // Moscow. Univ. Chem. Bull. 2008. N 6. P. 303. doi 10.3103/s0027131408060011
44. *Balaban A.T.* // J. Pure Appl. Chem. 1983. Vol. 55. N 2. P. 199. doi 10.1351/pac198855020199
45. *Станкевич М. И., Станкевич И. В., Зефирова Н. С.* // Усп. хим. 1988. Т. 58. № 3. С. 337.
46. *Xing R., Zhou B., Graovac A.* // Ars Combin. 2012. Vol. 104. P. 211.
47. *Vasilyev A., Stevanovi'c D.* // MATCH Commun. Math. Comput. Chem. 2014. Vol. 71. P. 657.
48. *Райхардт К.* Растворители и эффекты среды в органической химии. М.: Мир, 1991. 763 с.
49. *Боровиков В.* STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. СПб: Питер, 2003. 686 с.
50. *Паклин Н.Б., Орешков В.И.* Бизнес-аналитика: от данных к знаниям. СПб: Питер, 2009. 624 с.

Empirical Stability Models of Crown Ether Complexes with Alkaline and Alkaline Earth Metals in Some Pure Solvents

N. V. Bondarev*

V.N. Karazin Kharkov National University, Kharkov, 61022 Ukraine

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

Received February 4, 2020; revised February 4, 2020; accepted February 27, 2020

Based on exploratory and neural network methods for mathematical modeling of equilibria in solutions, an approach was developed to predict the stability constants of coronates by the properties of solvents, cations, and crown ethers. Exploration (factor, cluster, discriminant, canonical, decision trees), regression and neural network (with a teacher, Kohonen network) stability models of crown ether complexes (12C4, 13C4, 14C4, 15C4, 15C5, 18C6, 21C7, 24C8, B12C4, B15C5, CH15C5, CH18C6, DCH18C6, DCH21C7, DB18C6, DB21C7, DB24C8, DB27C9, DB30C10) with cations of alkali (Li^+ , Na^+ , K^+ , Cs^+ , Rb^+) and alkaline earth (Ca^{2+} , Sr^{2+} , Ba^{2+}) metals in aqueous and non-aqueous (acetone, acetonitrile, dimethyl sulfoxide, methanol, pyridine, dimethylformamide, dioxane, propylene carbonate, 1,2-dichloroethane, nitrobenzene) solutions were developed according to the properties of solvents (diameter of solvent molecule, Kamlet–Taft parameter, Dimroth–Reichardt parameter, dielectric constant), crown ethers (Balaban topological index) and cations (cation diameter) at 298.15 K.

Keywords: crown ethers, complexation constant, prospecting analysis, multiple linear regression, neural networks, modeling, forecasting