МАТЕМАТИЧЕСКИЕ МЕТОДЫ И ИНФОРМАЦИОННЫЕ СИСТЕМЫ В ХИМИЧЕСКОЙ ТЕХНОЛОГИИ
MATHEMATICS METHODS AND INFORMATION SYSTEMS IN CHEMICAL TECHNOLOGY

ISSN 2410-6593 (Print), ISSN 2686-7575 (Online)

https://doi.org/10.32362/2410-6593-2020-15-6-84-103

УДК 541.6

CC BY

НАУЧНАЯ СТАТЬЯ

Модели связи «структура-свойство» органических соединений на основе молекулярных графов с элементами пространственного строения молекул

Н.А. Шулаева, М.И. Скворцова[@], Н.А. Михайлова

МИРЭА – Российский технологический университет (Институт тонких химических технологий имени М.В. Ломоносова), Москва, 119571 Россия
[@]Автор для переписки, e-mail: skvorivan@mail.ru

Цели. *Цель работы – разработать, описать и протестировать общий алгоритмический метод построения моделей связи «структура-свойство» для органических соединений.*
Результаты. *Вышеуказанные модели строятся на основе статистического анализа данных по выборкам структур и свойств химических соединений и имеют вид корреляционных уравнений. Химические структуры в предложенном подходе представляются в виде специальных молекулярных графов с весами вершин и ребер, отражающих определенные особенности пространственного строения соответствующих молекул. Реализация метода происходит в два этапа. На первом этапе предполагается, что искомое уравнение связи «структура-свойство» имеет определенный аналитический вид и зависит от ряда подгоночных параметров и двух функций одной переменной, которые могут варьироваться. На этом этапе происходит отбор пары функций из заданного множества функций, дающих наилучшую модель. На втором этапе происходит модификация полученной наилучшей модели. Для этой цели первоначально проводится классификация вершин молекулярного графа по химическим символам соответствующих атомов и картинам их первого окружения; проводится также классификация ребер графа в соответствии с классами вершин, которые они соединяют. На основе полученной классификации вводятся числовые «поправки» к исходным весам вершин и ребер молекулярных графов, что позволяет улучшить модель, полученную на первом этапе. Конечным результатом процесса построения модели служит уравнение определенного вида с конкретными числовыми значениями всех его параметров. Приведены примеры применения предложенного метода для построения моделей связи «структура-свойство» для конкретных свойств и классов соединений, показывающие его эффективность. Рассматривались следующие физико-химические свойства и классы органических соединений: 1) температура кипения спиртов; 2) растворимость спиртов в воде; 3) температура кипения сульфидов; 4) индексы удерживания алкилфенолов.*

© Н.А. Шулаева, М.И. Скворцова, Н.А. Михайлова, 2020

Выводы. *Предложен общий алгоритмический метод построения корреляционных уравнений, связывающих структуру и свойства органических соединений. Приведены примеры его реализации. Метод может быть использован для любых классов органических соединений и любых их свойств, которые измеряются количественно. Модели, построенные на основе предложенного подхода, обладающие достаточно высоким качеством, могут быть использованы для расчета свойств соединений, для которых отсутствуют экспериментальные данные.*

Ключевые слова: *корреляции «структура-свойство», молекулярные графы, инварианты графа, компьютерная химия, математическая химия, длина химической связи, ковалентный радиус атома*

Для цитирования: Шулаева Н.А., Скворцова М.И., Михайлова Н.А. Модели связи «структура-свойство» органических соединений на основе молекулярных графов с элементами пространственного строения молекул. *Тонкие химические технологии.* 2020;15(6):84-103. https://doi.org/10.32362/2410-6593-2020-15-6-84-103

RESEARCH ARTICLE

Structure–property models of organic compounds based on molecular graphs with elements of the spatial structures of the molecules

Nadezhda A. Shulaeva, Mariya I. Skvortsova[@], Nataliya A. Mikhailova

MIREA – Russian Technological University (M.V. Lomonosov Institute of Fine Chemical Technologies), Moscow, 119571 Russia
[@]Corresponding author, e-mail: skvorivan@mail.ru

Objectives. *This article aims to describe, elaborate, and test a general algorithmic method for constructing the structure–property models for organic compounds.*
Results. *The construction of the models is based on the statistical analysis of some sets of chemical structures of definite classes with known property values. These models have some forms of correlation equations. For the representation of chemical structures in this method, the special weighted molecular graphs (MGs) that reflect some peculiarities of the spatial structures of the corresponding molecules are used. The proposed method is realized in two steps. First, it is assumed that the required structure–property equation has a definite form and depends on several adjusted numerical parameters and two changeable functions of one variable. In this step, from some set of functions, the pair of functions that provide the best model is selected. In the second step, the best model (from the previous step) is modified. For this purpose, the classification of the vertices of MG by the chemical symbols of the corresponding atoms and their first-order environments is fulfilled. Further, the graph edges are classified according to the classes of the vertices which they connect. Furthermore, the numerical correction terms for the initial weights of the vertices and edges are introduced, and they improve the obtained model. The final result of the model-construction process is the equation of the definite form containing concrete numerical values of its parameters. Some examples of the application of the elaborated method for constructing the structure–property models for the concrete properties and classes of compounds are presented. The following classes of organic compounds and their physicochemical properties are considered: 1) the boiling point of alcohols, 2) the water solubility of alcohols, 3) the boiling point of sulfides, and 4) the retention indices of alkylphenols. The obtained results indicate the efficiency of the proposed approach and the significance of introducing the second step to the method.*

Conclusions. *In this work, a general algorithmic and computerized method for constructing the structure–property models of organic compounds is suggested. Examples of the application of this method demonstrated its high efficiency. The method is suitable for any class of organic compounds and properties, which are quantitatively measured. Owing to its high efficiency, the structure–property models obtained by this approach can be employed to calculate the properties of chemical compounds for which experimental data are unavailable.*

Keywords: *structure–property correlations, weighted molecular graphs, graph invariants, computer chemistry, mathematical chemistry, length of the chemical bond, covalent atomic radius*

For citation: Shulaeva N.A., Skvortsova M.I., Mikhailova N.A. Structure–property models of organic compounds based on molecular graphs with elements of the spatial structures of the molecules. *Tonk. Khim. Tekhnol. = Fine Chem. Technol.* 2020;15(6):84-103 (Russ., Eng.). https://doi.org/10.32362/2410-6593-2020-15-6-84-103

ВВЕДЕНИЕ

Одной из важных задач математической и компьютерной химии является задача поиска количественных соотношений между структурой и свойствами химических соединений [1–11]. Найденные соотношения позволяют осуществлять прогноз свойств соединений (как реально существующих, так и гипотетических) по их структуре при помощи соответствующих расчетов, что может быть использовано для целенаправленного поиска соединений с заданным набором свойств. Следует отметить, что к настоящему времени синтезировано огромное количество химических веществ; однако экспериментальное определение их различных свойств с целью поиска соединений с нужными свойствами является во многих случаях весьма затруднительным с технической точки зрения и, кроме того, требует существенных финансовых и временных затрат. В связи с этим разработка различных математических методов моделирования связи между структурой и свойствами химических соединений является актуальной задачей.

Как правило, для построения моделей связи «структура-свойство» используется статистический подход, основанный на анализе некоторой заданной выборки химических структур с известными значениями изучаемого свойства. Для количественного описания структуры химических соединений прежде всего выбирается некоторый набор молекулярных параметров $x_1,...., x_n$; в качестве этих параметров могут быть использованы, например, какие-либо топологические, электронные или геометрические характеристики молекул. Далее предполагается, что свойство y связано с этими параметрами при помощи некоторой функции f: $y = f(x_1,...., x_n)$. Аналитический вид функции f задается исследователем (например, f – линейная или квадратичная функция), однако f зависит от ряда подбираемых параметров. Эти параметры определяются по известным данным для исходной (обучающей) выборки химических соединений так, чтобы уравнение $y = f(x_1,...., x_n)$ выполнялось бы как можно более точно (в некотором смысле) для исходного набора данных.

Для того чтобы оценить точность аппроксимации в построенной модели, обычно строится корреляция между расчетными и экспериментальными значениями изучаемого свойства для исходной (обучающей) выборки соединений. Для этой корреляции определяются коэффициент корреляции R, а также средняя относительная ошибка δ (в %), по которым делают выводы о качестве модели. Например, в [12] предложены следующие характеристики качества модели по величине коэффициента корреляции: $R \geq 0.990$ (outstanding), $R \geq 0.975$ (excellent), $R \geq 0.950$ (very good), $R \geq 0.925$ (good), $R \geq 0.900$ (fair). Заметим, что эти критерии, показывающие приемлемость модели, могут быть в ряде случаев выбраны и другими.

В исследованиях связи «структура-свойство» очень важны способы количественного описания молекулярных структур. Один из наиболее распространенных и естественных способов представления структуры молекулы – это изображение ее в виде некоторого графа с числовыми весами (или символьными метками); вершины и ребра такого графа соответствуют атомам и связям молекулы, а веса вершин и ребер количественно характеризуют особенности атомов и связей различных типов. В качестве топологических молекулярных дескрипторов $x_1,...., x_n$ в моделях связи «структура-свойство» используются какие-либо числовые инварианты этих графов [13–21].

Очевидно, что математические модели связи «структура-свойство», получаемые в рамках данного подхода, существенно зависят от выбора весов графов, представляющих химические структуры, так как вышеупомянутые инварианты графов определенным образом зависят от этих весов. Как правило, выбираемые веса вершин и ребер графа в конкретной задаче

не зависят ни от класса соединений, ни от рассматриваемого свойства. Например, для молекулярных графов алканов (эти графы строятся без учета атомов водорода) обычно полагают, что веса всех вершин равны нулю. В работе [15] приведены примеры весов вершин и весов ребер для взвешенных графов гетероатомных молекул; эти веса зависят от таких атомных характеристик как общее число электронов, число валентных электронов, число соседних атомов водорода, а также от параметров, характеризующих кратности химических связей. В литературе имеются также примеры весов вершин и ребер молекулярных графов, зависящих от ковалентных радиусов атомов, степеней вершин графа, расстояний между вершинами графа [22–24].

Следует отметить, что при построении моделей, связывающих структуру и свойства химических соединений, возникают вопросы о наилучшем выборе инвариантов графов x_1,..., x_n (молекулярных дескрипторов), аппроксимирующей функции f, а также весов вершин и ребер молекулярных графов, представляющих химические структуры. Наличие этих проблем обычно обусловлено отсутствием априорной информации о том, от каких особенностей структуры и каким образом зависит рассматриваемое свойство для данного класса соединений, а также бесконечным количеством вариантов выбора инвариантов графов, аппроксимирующей функции и весов графов.

Опишем теперь более подробно метод построения моделей связи «структура-свойство», основанный на оптимальном (в некотором смысле) подборе весов вершин взвешенных молекулярных графов, представляющих химические структуры, изложенный в работе [25]. Первоначально проводится некоторая классификация атомов, входящих в молекулярные структуры изучаемых соединений. Например, атомы могут быть разбиты на классы по химическим символам с учетом распределения типов связей; более детальная классификация атомов может быть получена, если для этой цели использовать их картины окружения первого порядка. Всем атомам одного, k-ого класса (k = 1,2,…) приписывается некоторый вес z_k (k = 1,2,…), представляющий собой число, которое на данном этапе неизвестно. Для дальнейших построений в каждой молекуле все атомы нумеруются. Затем предполагается, что зависимость изучаемого свойства y от структуры молекулы имеет следующий специальный вид (1):

$$y = \sum w_i w_j + c\,, \qquad (1)$$

где w_i, w_j – числовые веса атомов в молекуле с номерами i и j, определяемые тем классом, в который они попали при используемой классификации, т.е. $w_i = z_k$, если атом i оказался в k-ом классе; в выражении (1) сумма идет по всем связям (i,j) в молекуле, c – некоторая константа.

Далее неизвестные веса классов z_k (k = 1,2,…) подбираются так, чтобы соотношение (1) было бы как можно более точным на заданной выборке соединений. Для этой цели вводится нелинейная функция k переменных следующего вида (2):

$$F(z_1,...,z_k) = \sum\nolimits_p (y_p^{\text{эксп}} - y_p^{\text{расч}})^2\,, \qquad (2)$$

у которой ищется минимум и соответствующие ему значения переменных z_1,..., z_k. В формуле (2) $y_p^{\text{эксп}}$ – экспериментальное значение изучаемого свойства p-ого соединения, а $y_p^{\text{расч}}$ – аналитическое выражение для расчета свойства p-ого соединения, полученное на основе формулы (1) и зависящее от параметров z_1,..., z_k.

Таким образом, получаемое уравнение связи «структура-свойство» имеет вид уравнения (1), в котором известны параметры z_1,..., z_k, подобранные оптимальным образом (в вышеуказанном смысле). Уравнение (1) далее можно использовать для расчета значений свойств других соединений того же класса, не входящих в исходную выборку.

Основанием выбора формулы (1) для в вышеуказанных исследованиях послужили следующие факты. Один из наиболее широко используемых инвариантов молекулярных графов при моделировании связи «структура-свойство» – это индекс молекулярной связности (индекс Рандича) χ (3):

$$\chi = \sum \left(v_i v_j\right)^{-\frac{1}{2}}\,, \qquad (3)$$

где v_i, v_j – степени вершин i и j в графе, сумма идет по всем ребрам (i,j) графа [26]. Этот индекс обладает следующим свойством: для графов-деревьев с фиксированным числом вершин он принимает свои экстремальные значения (наибольшее и наименьшее) на наиболее и наименее разветвленных деревьях – цепи и графе-звезде. В силу этого свойства он может служить количественной мерой степени ветвления ациклической молекулы. Имеются обобщения индекса χ на случай более общих подграфов (например, цепочек разной длины; отметим, что индексу χ соответствует цепочка длины 1, т.е. ребро графа), а также на случай, когда в формуле для χ вместо степеней вершин стоят некоторые веса вершин, зависящие от характеристик соответствующих атомов. Вышеуказанные модификации индекса Рандича также нашли широкое применение в корреляциях «структура-свойство» [27, 28].

В работе [29] было предложено при построении корреляций «структура-свойство» заменить в формуле для χ величины v_i на величины вида $v_i + x$, где x – подбираемые параметры, зависящие от химических символов соответствующих атомов. На конкретном примере было показано, что

с таким «варьируемым» индексом связности можно получить более точную корреляцию вида «структура-свойство», чем с исходным индексом χ. Кроме того, в работе [29] было отмечено, что для каждого свойства и класса соединений оптимальным является свой набор подбираемых параметров.

Идея обобщить индекс Рандича путем введения подбираемых атомных параметров была реализована также в работах [30–35], в частности, для энтальпий испарения соединений таких классов, как алканы, спирты, эфиры и др. Соединения представлялись графами без учета атомов водорода. В этих работах рассматривался молекулярный дескриптор следующего вида (4):

$$^{0-3}\chi = {}^{0}\chi + \frac{^{1}\chi}{2} + \frac{^{2}\chi}{3} + \frac{^{3}\chi}{4}, \quad (4)$$

где

$$^{0}\chi = \sum\left(\ln\delta_i\right)^{-1}, \; ^{1}\chi = \sum\left(\ln\left(\delta_i\delta_j\right)\right)^{-1},$$
$$^{2}\chi = \sum\left(\ln\left(\delta_i\delta_j\delta_k\right)\right)^{-1}, \; ^{3}\chi = \sum\left(\ln\left(\delta_i\delta_j\delta_k\delta_l\right)\right)^{-1}. \quad (5)$$

В формуле (5) для $^{0}\chi$ суммирование идет по всем неводородным атомам молекулы, занумерованных числами i = 1,2,...,n. В формулах для $^{1}\chi$, $^{2}\chi$, $^{3}\chi$ суммирование идет по всем цепочкам атомов, содержащих 1, 2, 3 последовательные связи, с номерами атомов в этих цепочках *i,j* или *i,j,k* или *i,j,k,l*, соответственно; числа вида δ_i – подбираемые параметры атомов, зависящие от способа их классификации. Далее, предполагалось, что зависимость рассматриваемого свойства от параметра $^{0-3}\chi$ имеет следующий вид (6):

$$y = a \times {}^{0-3}\chi + b, \quad (6)$$

где константы a и b подбираются по методу наименьших квадратов.

В работе [36] также рассматривается некоторая модификация классического индекса Рандича, основанная на подборе атомных параметров. При этом часть параметров, отнесенных к атомам углерода, фиксировалась (они принимались равными степеням вершин в соответствующем графе), а другая часть подбиралась.

В настоящей работе предложен новый общий метод установления количественных соотношений между структурой и свойствами органических соединений, разработанный в рамках статистического подхода к решению таких задач. Кроме того, проведено его тестирование. Метод основан на представлении структур изучаемых соединений в виде взвешенных молекулярных графов, веса вершин и ребер которых отражают элементы пространственного строения соответствующих молекул, с последующей корректировкой этих весов. Предлагаемый подход является обобщением и развитием подхода, изложенного ранее в работе [25].

ОПИСАНИЕ МЕТОДА ПОСТРОЕНИЯ МОДЕЛЕЙ СВЯЗИ «СТРУКТУРА-СВОЙСТВО»

Построение взвешенных молекулярных графов органических соединений

Пусть имеется некоторая выборка структурных формул органических соединений определенного класса, с известными численными значениями некоторого физико-химического свойства, полученными экспериментально и приведенными в литературе.

Построение любых моделей связи «структура-свойство» в рамках вышеописанного статистического подхода предполагает предварительное построение специальных молекулярных графов (МГ) рассматриваемых соединений. В данной работе рассматриваются взвешенные МГ, вершины которых соответствуют неводородным атомам молекулы, а ребра – химическим связям между этими атомами. Предполагается, что вершины МГ занумерованы числами 1,2,...,n, а ребро, образованное парой вершин с номерами i и j ($i < j$), обозначено символом (i,j). Веса вершин и ребер МГ отражают элементы пространственного строения молекулы, а именно, веса вершин w_{ii} берутся равными ковалентным радиусам соответствующих атомов, а веса ребер w_{ij} равны длинам соответствующих связей. Используемые в работе значения радиусов атомов и длин связей приведены в табл. 1 и 2 ([23] и [37], соответственно).

Таблица 1. Ковалентные радиусы атомов (Å) [23]
Table 1. Covalent radii of the atoms (Å) [23]

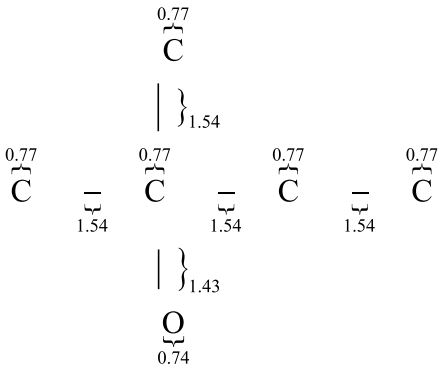
№	Атом Atom	Радиус атома, Å Atom radius, Å
1	Csp^3	0.77
2	Csp^2	0.67
3	Csp	0.60
4	Nsp^3	0.74
5	Nsp^2	0.62
6	Nsp	0.55
7	Osp^3	0.74
8	Osp^2	0.62
9	F	0.72
10	Psp^3	0.10
11	Psp^2	1.00
12	Ssp^3	1.04
13	Ssp^2	0.94
14	Cl	0.99
15	Br	1.14
16	I	1.33

Таблица 2. Средние длины связей в органических молекулах (Å) [37]
Table 2. Average bond lengths in the organic molecules (Å) [37]

№	Связь Bond	Длина связи, Å Bond length, Å	№	Связь Bond	Длина связи, Å Bond length, Å
1	C4–C4	1.54	23	C2–C1	1.63
2	C4–C3	1.52	24	C3=C3	1.34
3	C4–C2	1.46	25	C3=C2	1.31
4	C4–N3	1.47	26	C3=N2	1.32
5	C4–N2	1.47	27	C3=O1	1.22
6	C4–O2	1.43	28	C2=C2	1.28
7	C4–S2	1.81	29	C2=N2	1.32
8	C4–F	1.40	30	C2=O1	1.16
9	C4–Cl	1.76	31	N3=O1	1.24
10	C4–Br	1.94	32	N2=O1	1.22
11	C4–I	2.14	33	N2=N2	1.25
12	C3–C3	1.46	34	C2≡C2	1.20
13	C3–C2	1.45	35	C2≡N1	1.16
14	C3–N3[1]	1.40	36	C3=C3[2]	1.40
15	C3–N2	1.40	37	C3=N2[2]	1.34
16	C3–O2	1.36	38	C3–F[2]	1.32
17	C3–F	1.33	39	C3–Cl[2]	1.71
18	C3–Cl	1.73	40	C3–Br[2]	1.89
19	C2–C2	1.38	41	C3–I[2]	2.10
20	C2–N3	1.33	42	C3–O2[2]	1.36
21	C2–O2	1.33	43	C3–N3[2]	1.48
22	C2–F	1.30	44	C3–S2[2]	1.81

В табл. 2 число после символа атома – это число соседних атомов для данного атома в молекуле; верхний индекс 1 в № 14 означает, что длина связи для N–C=O группы равна 1.32; верхний индекс 2 в №№ 36–44 означает, что C3 – это атом углерода в бензольном кольце.

Пример описанного выше взвешенного МГ, соответствующего структурной формуле 2-метил-2-бутанола, приведен на рисунке (вершины МГ не занумерованы).



Взвешенный молекулярный граф 2-метил-2-бутанола.
Weighted molecular graph of 2-methyl-2-butanol.

Описание первоначальной модели связи «структура-свойство» и методики подбора ее параметров

В предлагаемом подходе к моделированию связи «структура-свойство» первоначально предполагается, что соответствующее уравнение, связывающее свойство *y* и некоторые структурные характеристики молекулы (или ее МГ), имеет следующий вид (7):

$$y = a\sum f_1\left(w_{ij}\right) + b\sum f_2\left(w_{ii}\times w_{jj}\right) + c\,, \qquad (7)$$

где w_{ij} – вес ребра (i,j) в МГ, w_{ii} – вес вершины i в МГ, суммирование в каждой сумме идет по всем ребрам (i,j) в МГ. Функции $f_1(x)$ и $f_2(x)$ – некоторые фиксированные функции одной переменной. Обе функции выбираются независимо из какого-либо заданного множества функций. В данной работе рассматривается следующее множество функций: $f(x) = \ln x$, $f(x) = x^k$, где k принимает значения ±1, ±0.5 (т.е. рассматривается 5 функций). Заметим, что это множество функций можно расширить или изменить. Таким образом, модель определяется заданием пары функций $f_1(x)$ и $f_2(x)$. Параметры a, b, c подбираются по методу наименьших квадратов по

заданной (обучающей) выборке соединений для каждой пары функций $f_1(x)$ и $f_2(x)$. Итак, в данном случае можно построить 25 разных моделей.

Для оценки точности каждой из этих 25 моделей (т.е. ее качества) на обучающей выборке соединений рассчитываются коэффициент корреляции R и среднеквадратичное отклонение s для корреляции между экспериментальными и расчетными значениями изучаемого свойства. Из полученного множества моделей на основе параметров R и s выбирается одна, наиболее точная модель. Эта модель далее может быть усовершенствована.

Заметим, что для оценки точности модели на исходной выборке соединений можно также использовать среднюю и максимальную абсолютную (или относительную) ошибки расчета величины свойства ($\Delta_{ср}$, $\Delta_{макс}$, $\delta_{ср}$, $\delta_{макс}$, соответственно).

Корректировка модели, основанная на модификации весов вершин и ребер молекулярных графов

Для улучшения полученной модели (с выбранной парой функций $f_1(x)$ и $f_2(x)$) используется корректировка весов вершин и ребер МГ. Эта корректировка проводится следующим образом. Первоначально вершины и ребра МГ по определенным критериям разбиваются на некоторые классы.

Классификация вершин МГ проводится по химическим символам соответствующих атомов и по картинам их окружения первого порядка. Классы вершин нумеруются произвольным образом числами 1,2,…,n_1. Классу с номером p первоначально приписывается некоторый неопределенный числовой «вес» x_p (p = 1,2,….,n_1). Связи классифицируются в соответствии с классами атомов, которые они соединяют; классы связей также нумеруются произвольным образом числами 1,2,…,n_2. Классу с номером q первоначально приписывается некоторый неопределенный числовой «вес» z_q (q = 1,2,….,n_2).

Далее, вместо исходных весов вершин w_{ii} рассматриваются веса вида $w_{ii} + x_p$, где p – номер класса, к которому принадлежит i-ый атом, а вместо исходных весов ребер w_{ij} – аналогично подкорректированные веса, имеющие вид $w_{ij} + z_q$, где q – номер класса, к которому принадлежит ребро (i,j).

Далее эти «поправки», а также коэффициенты a, b и c (которые предполагаются неизвестными) подбираются одновременно путем минимизации функции многих переменных вида (8):

$$F\left(a,b,c,x_1,x_2,x_3,...,z_1,z_2,z_3,...\right) = \\ = \sum_k \left(y_{эксп,k} - y_{расч,k}\right)^2, \quad (8)$$

где $y_{эксп,k}$ – экспериментальное значение свойства k-го соединения, $y_{расч,k}$ – это расчетное значение свойства этого соединения, полученное по формуле вида (7), где вместо исходных весов w_{ij} и w_{ii} стоят их подкорректированные выражения; суммирование идет по всем соединениям обучающей выборки.

Для оптимального подбора параметров a, b, c, x_1, x_2, x_3,… в этой задаче может быть использована надстройка «Поиск решения» программы Microsoft Excel.

Отметим, что для уточнения модели может быть проведена только корректировка весов вершин; при этом веса ребер остаются неизменными.

Для оценки качества модели на обучающей выборке соединений определяются коэффициент корреляции R и среднеквадратичное отклонение s для корреляции между экспериментальными и расчетными значениями изучаемого свойства, а также средняя и максимальная абсолютная (или относительная) ошибка расчета величины свойства ($\Delta_{ср}$, $\Delta_{макс}$, $\delta_{ср}$, $\delta_{макс}$, соответственно).

Для оценки прогнозирующей способности модели используется некоторый новый набор химических соединений того же класса с известными значениями рассматриваемого свойства (тестовая выборка). Для этих соединений по полученной формуле рассчитываются значения их свойств и определяются коэффициент корреляции R_1 и среднеквадратичное отклонение s_1 для корреляции между экспериментальными и расчетными значениями изучаемого свойства, а также параметры $\Delta_{ср,1}$, $\Delta_{макс,1}$, $\delta_{ср,1}$, $\delta_{макс,1}$, аналогичные вышеописанным параметрам для обучающей выборки.

ПРИМЕРЫ ПОСТРОЕНИЯ МОДЕЛЕЙ СВЯЗИ «СТРУКТУРА-СВОЙСТВО» НА ОСНОВЕ ПРЕДЛОЖЕННОГО МЕТОДА

В этом разделе приведены примеры применения разработанного метода для построения моделей связи «структура-свойство» для ряда физико-химических свойств органических соединений некоторых классов, а также дан анализ полученных результатов.

В каждом примере рассматривается некоторая выборка соединений, разделенная на обучающую и тестовую выборки; N – число соединений в исходной выборке; N_1 и N_2 – число соединений в обучающей и тестовой выборках, соответственно. Исходные данные представлены в виде таблицы, содержащей названия соединений и значения их рассматриваемого физико-химического свойства, взятые из литературы; соединения, входящие в тестовую выборку, отмечены значком *.

В соответствии с описанным выше методом в каждом примере на первом этапе исследований строятся взвешенные МГ рассматриваемых соединений. Затем для обучающей выборки для каждой пары выбранных функций $f_1(x)$ и $f_2(x)$ вычисляются неизвестные константы a, b, c по методу наименьших квадратов, а также параметры R и s. Из полученных моделей

по параметрам R и s выбирается наилучшая модель, т.е. выбирается наилучшая пара функций $f_1(x)$ и $f_2(x)$. Для этой модели определяются параметры R, s, $\delta_{ср}$, $\delta_{макс}$, характеризующие качество модели на обучающей выборке, а также аналогичные параметры R_1, s_1, $\delta_{ср,1}$, $\delta_{макс,1}$, характеризующие качество модели на тестовой выборке.

На втором этапе исследований происходит уточнение полученной наилучшей модели. Для этой цели первоначально проводится классификация вершин МГ по химическим символам соответствующих атомов и картинам первого окружения. Результаты классификации сводятся в таблицу, содержащую структурный фрагмент, определяющий класс, а также описание класса.

Далее, на основе полученной классификации вершин вводятся поправки x_1, x_2,… к исходным весам вершин. Найденные значения поправок x_1, x_2,… и коэффициентов a, b, c для улучшенной модели сводятся в таблицу. Конечным результатом процесса построения модели служит уравнение определенного вида с конкретными числовыми значениями всех его параметров, позволяющее рассчитать значение изучаемого свойства для любого соединения заданного класса. Для этого уравнения также определяются параметры R, s, $\delta_{ср}$, $\delta_{макс}$, R_1, s_1, $\delta_{ср,1}$, $\delta_{макс,1}$, характеризующие качество модели на обучающей и тестовой выборке, соответственно.

Наконец, на третьем этапе исследований проводится анализ полученных результатов. С этой целью все статистические характеристики для всех четырех случаев (исходная/уточненная модель; обучающая/тестовая выборка) сводятся в одну таблицу; затем проводится сравнение этих характеристик для исходной и уточненной модели, на основе которого делается заключение о целесообразности введения поправок к весам вершин МГ для улучшения модели. Кроме того, дается оценка качества итоговой, уточненной модели.

а) Температура кипения спиртов

В качестве исходных данных для построения модели была рассмотрена выборка спиртов с их значениями температур кипения, $N = 31$, $N_1 = 21$, $N_2 = 10$ [29]. Исходные данные приведены в табл. 3.

Наилучшей моделью, отобранной на первом этапе, является уравнение вида (9):

$$y = a\sum\left(w_{ij}\right)^{-1} + b\sum\left(w_{ii} \times w_{jj}\right)^{-\frac{1}{2}} + c\,. \qquad (9)$$

Далее проводится классификация вершин МГ; результаты классификации приведены в табл. 4.

Таблица 3. Спирты и их температуры кипения [29]
Table 3. Alcohols and their boiling points [29]

№	Название соединения Compound	Температура кипения, °С Boiling point, °С
1	Этанол / Ethanol	78.0
2	Пропанол / Propanol	97.1
3	2-Пропанол / 2-Propanol	82.4
4	Бутанол* / Butanol*	117.6
5	2-Метил-1-пропанол / 2-Methyl-1-propanol	108.1
6	2-Бутанол* / 2-Butanol*	99.5
7	2-Метил-2-пропанол / 2-Methyl-2-propanol	82.4
8	Пентанол* / Pentanol*	138.0
9	3-Метил-1-бутанол / 3-Methyl-1-butanol	131.0
10	2-Метил-1-бутанол / 2-Methyl-1-butanol	128.0
11	2-Пентанол* / 2-Pentanol*	119.3
12	3-Пентанол / 3-Pentanol	116.2
13	3-Метил-2-бутанол* / 3-Methyl-2-butanol*	112.9
14	2-Метил-2-бутанол / 2-Methyl-2-butanol	102.3
15	Гексанол / Hexanol	157.6
16	3-Метил-1-пентанол / 3-Methyl-1-pentanol	153.0
17	4-Метил-1-пентанол / 4-Methyl-1-pentanol	151.9
18	2-Метил-1-пентанол* / 2-Methyl-1-pentanol*	149.0
19	2-Этил-1-бутанол / 2-Ethyl-1-butanol	147.0
20	2,3-Диметил-1-бутанол* / 2,3-Dimethyl-1-butanol*	144.5
21	3,3-Диметил-1-бутанол / 3,3-Dimethyl-1-butanol	143.0

Таблица 3. Окончание
Table 3. Continued

№	Название соединения Compound	Температура кипения, °C Boiling point, °C
22	2-Гексанол* / 2-Hexanol*	140.0
23	2,2-Диметил-1-бутанол / 2,2-Dimethyl-1-butanol	136.5
24	3-Гексанол / 3-Hexanol	135.0
25	3-Метил-2-пентанол* / 3-Methyl-2-pentanol*	134.3
26	4-Метил-2-пентанол / 4-Methyl-2-pentanol	131.6
27	2-Метил-3-пентанол* / 2-Methyl-3-pentanol*	126.5
28	3-Метил-3-пентанол / 3-Methyl-3-pentanol	122.4
29	2-Метил-2-пентанол / 2-Methyl-2-pentanol	121.1
30	3,3-Диметил-2-бутанол / 3,3-Dimethyl-2-butanol	120.4
31	2,3-Диметил-2-бутанол / 2,3-Dimethyl-2-butanol	118.4

Таблица 4. Классификация вершин МГ, соответствующих молекулам спиртов
Table 4. Classification of vertices of MG, corresponding to alcohol molecules

№	Структурный фрагмент, определяющий класс атомов Structural fragment that defines the class of atoms	Описание класса атомов Description of the atom class
1	$\mathbf{C_1}$— C	Центральный атом C_1 связан с одним атомом C The central C_1 atom is bonded to one C atom
2	C— $\mathbf{C_2}$— C	Центральный атом C_2 связан с двумя атомами C The central C_2 atom is bonded to two C atoms
3	C— $\mathbf{C_3}$— C | C	Центральный атом C_3 связан с тремя атомами C The central C_3 atom is bonded to three C atoms
4	C | C — $\mathbf{C_4}$ — C | C	Центральный атом C_4 связан с четырьмя атомами C The central C_4 atom is bonded to four C atoms
5	C — $\mathbf{C_5}$ — O	Центральный атом C_5 связан с одним атомом C и с одним атомом O The central C_5 atom is bonded to one C and one O atoms
6	C— $\mathbf{C_6}$ — C | O	Центральный атом C_6 связан с двумя атомами C и с одним атомом O The central C_6 atom is bonded to two C and one O atoms
7	C | C — $\mathbf{C_7}$ — C | O	Центральный атом C_7 связан с тремя атомами C и с одним атомом O The central C_7 atom is bonded to three C and one O atoms
8	C — **O**	Центральный атом O связан с одним атомом C The central O atom is bonded to one C atom

На основе этой классификации вводятся поправки x_1–x_8 (коррекционные параметры) к весам вершин МГ. Найденные поправки x_1, x_2, …, x_8 и коэффициенты a, b, c приведены в табл. 5.

Таким образом, улучшенная модель имеет следующий вид (10):

$$y = a\sum\left(w_{ij}\right)^{-1} + \\ +b\sum\left(\left(w_{ii}+x_n\right)\times\left(w_{jj}+x_m\right)\right)^{-1/2} + c, \quad (10)$$

где n и m – это номера классов вершин, к которым принадлежат соответствующие вершины i и j; значения

поправок x_n и x_m, а также коэффициентов a, b, c берутся из табл. 5; суммирование в каждой сумме идет по всем ребрам (i,j) в МГ.

В табл. 6 приведены некоторые статистические параметры построенных моделей для всех четырех случаев (исходная/уточненная модель; обучающая/тестовая выборка).

Из табл. 6 следует, что введение поправок к весам вершин МГ значительно улучшает исходную модель (по всем показателям). Таким образом, уравнение вида (10) более точно, чем уравнение (9), отражает связь между структурой спиртов и их температурой кипения.

б) Растворимость спиртов в воде

В качестве исходных данных для построения модели была рассмотрена выборка спиртов с их значениями растворимости в воде, $-\log X$, где X – мольная доля вещества в растворе, $N = 30$, $N_1 = 20$, $N_2 = 10$ [38]. Исходные данные приведены в табл. 7.

Таблица 5. Значения поправок x_1–x_8 и коэффициентов a, b, c для улучшенной модели
Table 5. Values of the correction parameters, x_1–x_8, and the coefficients a, b, and c for the improved model

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	a	b	c
0.0348	0.0003	0.0112	0.0170	0.0001	0.0117	0.0174	0.00303	15898	7936.4	−549.4

Таблица 6. Некоторые статистические параметры исходной/модифицированной модели для обучающей/тестовой выборки для температуры кипения спиртов
Table 6. Some statistical parameters of the initial/modified models for the training/test samples for the boiling point of alcohols

Статистические параметры модели Statistical parameters of the model	Исходная модель (без поправок к весам вершин МГ) Initial model (without the corrections of w_{ii} of MG)		Модифицированная модель (с поправками к весам вершин МГ) Modified model (with the corrections of w_{ii} of MG)	
	Обучающая выборка Training sample	Тестовая выборка Test sample	Обучающая выборка Training sample	Тестовая выборка Test sample
s	13.19	10.09	1.82	2.04
R	0.838	0.796	0.997	0.992
$\delta_{ср}$ / δ_{avg}	9.23%	6.18%	1.21%	1.39%
$\delta_{макс}$ / δ_{max}	25.43%	13.25%	3.96%	3.05%

Наилучшей моделью, отобранной на первом этапе, является уравнение вида (11):

$$y = a\sum\left(w_{ij}\right)^{1/2} + b\sum\left(w_{ii}\times w_{jj}\right)^{-1/2} + c\,. \qquad (11)$$

Далее проводится классификация вершин МГ; результаты классификации приведены в табл. 4. На основе этой классификации вводятся поправки x_1–x_8 (коррекционные параметры) к весам вершин МГ. Найденные поправки x_1, x_2, …, x_8 и коэффициенты a, b, c приведены в табл. 8.

Следовательно, улучшенная модель имеет следующий вид (12):

$$y = a\sum\left(w_{ij}\right)^{1/2} + b\sum\left(\left(w_{ii}+x_n\right)\times\left(w_{jj}+x_m\right)\right)^{-1/2} + c, \qquad (12)$$

где n и m – это номера классов вершин, к которым принадлежат соответствующие вершины i и j; значения поправок x_n и x_m, а также коэффициентов a, b, c берутся из табл. 8; суммирование в каждой сумме идет по всем ребрам (i,j) в МГ.

В табл. 9 приведены некоторые статистические параметры построенных моделей для всех четырех случаев (исходная/уточненная модель; обучающая/тестовая выборка).

Из табл. 9 следует, что введение поправок к весам вершин МГ значительно улучшает исходную модель (по всем показателям). Таким образом, уравнение вида (12) более точно, чем уравнение (11), отражает связь между структурой спиртов и их растворимостью в воде.

С целью дальнейшего улучшения модели проведем теперь классификацию ребер МГ (или связей в структурной формуле), основанную на указанной выше классификации вершин МГ. Были выделены следующие классы ребер:

1) C_1–C_2, 2) C_1–C_3, 3) C_1–C_4, 4) C_1–C_5, 5) C_1–C_6, 6) C_1–C_7, 7) C_2–C_2, 8) C_2–C_3, 9) C_2–C_4, 10) C_2–C_5, 11) C_2–C_6, 12) C_2–C_7, 13) C_3–C_5, 14) C_3–C_6, 15) C_3–C_7, 16) C_4–C_5, 17) C_4–C_6, 18) O–C_5, 19) O–C_6, 20) O–C_7.

Таблица 7. Спирты и их растворимости в воде [38]
Table 7. Alcohols and their water-solubility values [38]

№	Название соединения Compound	Растворимость в воде, −log*X* Solubility in water, −log*X*
1	1-Бутанол / 1-Butanol	1.750
2	2-Метил-1-пропанол / 2-Methyl-1-propanol	1.743
3	2-Бутанол* / 2-Butanol*	1.724
4	1-Пентанол / 1-Pentanol	2.332
5	3-Метил-1-бутанол / 3-Methyl-1-butanol	2.254
6	2-Метил-1-бутанол* / 2-Methyl-1-butanol*	2.207
7	2-Пентанол / 2-Pentanol	2.025
8	3-Пентанол* / 3-Pentanol*	1.961
9	3-Метил-2-бутанол / 3-Methyl-2-butanol	1.926
10	2-Метил-2-бутанол / 2-Methyl-2-butanol	1.608
11	2,2-Диметил-1-пропанол / 2,2-Dimethyl-1-propanol	2.030
12	1-Гексанол* / 1-Hexanol*	2.957
13	2-Гексанол / 2-Hexanol	2.612
14	3-Гексанол / 3-Hexanol	2.542
15	3-Метил-3-пентанол* / 3-Methyl-3-pentanol*	2.109
16	2-Метил-2-пентанол / 2-Methyl-2-pentanol	2.233
17	2-Метил-3-пентанол* / 2-Methyl-3-pentanol*	2.445
18	3-Метил-2-пентанол* / 3-Methyl-2-pentanol*	2.458
19	2,3-Диметил-2-бутанол / 2,3-Dimethyl-2-butanol	2.118
20	3,3-Диметил-1-бутанол / 3,3-Dimethyl-1-butanol	2.870
21	3,3-Диметил-2-бутанол / 3,3-Dimethyl-2-butanol	2.359
22	4-Метил-1-пентанол / 4-Methyl-1-pentanol	2.737
23	4-Метил-2-пентанол* / 4-Methyl-2-pentanol*	2.534
24	2-Этил-1-бутанол / 2-Ethyl-1-butanol	2.956
25	1-Гептанол / 1-Heptanol	3.554
26	2-Метил-2-гексанол / 2-Methyl-2-hexanol	2.820
27	3-Метил-3-гексанол* / 3-Methyl-3-hexanol*	2.729
28	3-Этил-3-пентанол / 3-Ethyl-3-pentanol	2.579
29	2,3-Диметил-2-пентанол / 2,3-Dimethyl-2-pentanol	2.615
30	2,3-Диметил-3-пентанол* / 2,3-Dimethyl-3-pentanol*	2.588

Таблица 8. Значения поправок x_1–x_8 и коэффициентов *a*, *b*, *c* для улучшенной модели
Table 8. Values of the correction parameters, x_1–x_8, and the coefficients *a*, *b*, and *c* for the improved model

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	*a*	*b*	*c*
0.0005	0.0011	0.0011	0.0011	0.0001	0.0002	−0.0003	−0.0002	213.25	−203.61	15.14

В соответствии с этой классификацией были введены и найдены поправки z_1–z_{20} к весам ребер, а также одновременно найдены и новые значения параметров *a*, *b*, *c*, x_1–x_8. В табл. 10 приведены некоторые статистические параметры построенных модифицированных моделей для всех четырех случаев (модель с поправками к весам вершин/к весам вершин и ребер; обучающая/тестовая выборка).

Следует отметить, что введение дополнительных параметров в корреляционное уравнение привело к улучшению всех показателей модели на обучающей выборке. Однако на тестовой выборке модель стала несколько хуже. В связи с этим проведение дополнительной классификации ребер МГ не представляется целесообразным.

в) Температура кипения сульфидов

В качестве исходных данных для построения модели была рассмотрена выборка сульфидов с их значениями температур кипения, $N = 30$, $N_1 = 20$, $N_2 = 10$ [39]. Исходные данные приведены в табл. 11.

Таблица 9. Некоторые статистические параметры исходной/модифицированной модели для обучающей/тестовой выборки для растворимостей спиртов в воде
Table 9. Some statistical parameters of the initial/modified models for the training/test samples for the water solubility of alcohols

Статистические параметры модели Statistical parameters of the model	Исходная модель (без поправок к весам вершин МГ) Initial model (without the corrections of w_{ii} of MG)		Модифицированная модель (с поправками к весам вершин МГ) Modified model (with the corrections of w_{ii} of MG)	
	Обучающая выборка Training sample	Тестовая выборка Test sample	Обучающая выборка Training sample	Тестовая выборка Test sample
s	0.301	0.239	0.078	0.091
R	0.794	0.798	0.988	0.973
$\delta_{ср}$ / δ_{avg}	9.25%	6.31%	2.40%	3.43%
$\delta_{макс}$ / δ_{max}	28.53%	17.49%	6.29%	17.40%

Таблица 10. Некоторые статистические параметры модифицированных моделей с поправками к весам вершин и с поправками к весам вершин и ребер для обучающей и тестовой выборок для растворимости спиртов в воде
Table 10. Some statistical parameters of the modified models with the correction parameters to vertex weights and to weights of vertices and edges for the training and test samples for the water solubility of alcohols

Статистические параметры модели Statistical parameters of the model	Модифицированная модель (с поправками к весам вершин МГ) Modified model (with corrections to the vertex weights of MG)		Модифицированная модель (с поправками к весам вершин и ребер МГ) Modified model (with corrections to the weights of vertices and edges of MG)	
	Обучающая выборка Training sample	Тестовая выборка Test sample	Обучающая выборка Training sample	Тестовая выборка Test sample
s	0.078	0.091	0.026	0.099
R	0.988	0.973	0.999	0.968
$\delta_{ср}$ / δ_{avg}	2.40%	3.43%	0.69%	3.43%
$\delta_{макс}$ / δ_{max}	6.29%	17.40%	2.48%	18.86%

Таблица 11. Сульфиды и их температуры кипения [39]
Table 11. Sulfides and their boiling points [39]

№	Название соединения Compound	Температура кипения, °C Boiling point, °C
1	Метилэтилсульфид / Methyl ethyl sulfide	66.6
2	Метилпропилсульфид / Methyl propyl sulfide	95.5
3	Диэтилсульфид / Diethyl sulfide	92.0
4	Метилизопропилсульфид / Methylisopropyl sulfide	84.4
5	Этилизопропилсульфид / Ethylisopropyl sulfide	107.4
6	Метилбутилсульфид* / Methyl butyl sulfide*	123.2
7	Метилизобутилсульфид / Methyl isobutyl sulfide	112.5
8	Этилпропилсульфид* / Ethylpropyl sulfide*	118.5
9	Метил-*трет*-бутилсульфид / Methyl *tert*-butyl sulfide	101.5
10	Метиламилсульфид / Methylamyl sulfide	145.0
11	Этилбутилсульфид / Ethyl butyl sulfide	144.2
12	Дипропилсульфид* / Dipropyl sulfide*	142.8
13	Пропилизопропилсульфид / Propylisopropyl sulfide	132.0
14	Этилизобутилсульфид / Ethyl isobutyl sulfide	134.2
15	Метилизоамилсульфид / Methylisoamyl sulfide	137.0

Таблица 11. Окончание
Table 11. Continued

№	Название соединения Compound	Температура кипения, °С Boiling point, °С
16	Метил-2-метилбутилсульфид* / Methyl 2-methylbutyl sulfide*	139.0
17	Этил-*втор*-бутилсульфид / Ethyl *sec*-butyl sulfide	133.6
18	Этил-*трет*-бутилсульфид* / Ethyl *tert*-butyl sulfide*	120.4
19	Диизопропилсульфид / Diisopropyl sulfide	120.0
20	Метил-1-этилпропилсульфид / Methyl 1-ethylpropyl sulfide	137.0
21	Метил-*втор*-бутилсульфид * / Methyl *sec*-butyl sulfide*	114.5
22	Метил-*трет*-амилсульфид / Methyl *tert*-amyl sulfide	128.3
23	Метил-1,2-диметилпропилсульфид / Methyl-1,2-dimethylpropyl sulfide	133.0
24	Метилгексилсульфид* / Methylhexyl sulfide*	171.0
25	Пропилбутилсульфид / Propyl butyl sulfide	166.0
26	Пропилизобутилсульфид / Propyl isobutyl sulfide	155.0
27	Изопропилизобутилсульфид* / Isopropyl isobutyl sulfide*	145.0
28	Этил-2-метилбутилсульфид* / Ethyl 2-methylbutyl sulfide*	159.0
29	Пропил-*трет*-бутилсульфид / Propyl *tert*-butyl sulfide	138.0
30	Изопропил-*втор*-бутилсульфид* / Isopropyl *sec*-butyl sulfide*	142.0

Наилучшей моделью, отобранной на первом этапе, является уравнение следующего вида (13):

$$y = a\sum\left(w_{ij}\right)^{-1} + b\sum\ln\left(w_{ii}\times w_{jj}\right) + c\,. \qquad (13)$$

Далее проводится классификация вершин МГ; результаты классификации приведены в табл. 12.

На основе этой классификации вводятся поправки x_1–x_8 к весам вершин МГ. Найденные поправки x_1, x_2, …, x_8 и коэффициенты a, b, c приведены в табл. 13.

Таким образом, наилучшая модель имеет следующий вид (14):

$$y = a\sum\left(w_{ij}\right)^{-1} + \\ +b\sum\ln\left(\left(w_{ii}+x_n\right)\times\left(w_{jj}+x_m\right)\right)+c, \qquad (14)$$

где n и m – это номера классов вершин, к которым принадлежат соответствующие вершины i и j; значения поправок x_n и x_m, а также коэффициентов a, b, c берутся из

Таблица 12. Классификация вершин МГ, соответствующих молекулам сульфидов
Table 12. Classification of vertices of MG, corresponding to sulfide molecules

№	Структурный фрагмент, определяющий класс атомов Structural fragment that defines the class of atoms	Описание класса атомов Description of the atom class
1	$\mathbf{C_1}$— C	Центральный атом C_1 связан с одним атомом C The central C_1 atom is bonded to one C atom
2	C— $\mathbf{C_2}$— C	Центральный атом C_2 связан с двумя атомами C The central C_2 atom is bonded to two C atoms
3	C— $\mathbf{C_3}$— C | C	Центральный атом C_3 связан с тремя атомами C The central C_3 atom is bonded to three C atoms
4	$\mathbf{C_4}$— S	Центральный атом C_4 связан с одним атомом S The central C_4 atom is bonded to one S atom
5	C — $\mathbf{C_5}$ — S	Центральный атом C_5 связан с одним атомом C и с одним атомом S The central C_5 atom is bonded to one C and one S atoms
6	C — $\mathbf{C_6}$ — S | C	Центральный атом C_6 связан с двумя атомами C и с одним атомом S The central C_6 atom is bonded to two C and one S atoms

Таблица 12. Окончание
Table 12. Continued

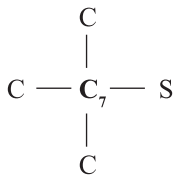
№	Структурный фрагмент, определяющий класс атомов Structural fragment that defines the class of atoms	Описание класса атомов Description of the atom class
7		Центральный атом C_7 связан с тремя атомами C и с одним атомом S The central C_7 atom is bonded to three C and one S atoms
8		Центральный атом S связан с двумя атомами C The central S atom is bonded to two C atoms

Таблица 13. Значения поправок x_1–x_8 и коэффициентов a, b, c для улучшенной модели
Table 13. Values of the correction parameters, x_1–x_8, and the coefficients a, b, and c for the improved model

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	a	b	c
0.00016	0.00014	0.00002	0.00017	0.00006	0.00006	0.00021	0.00006	7284.7	9091.2	4055.93

табл. 13; суммирование в каждой сумме идет по всем ребрам (i,j) в МГ.

В табл. 14 приведены некоторые статистические параметры моделей для всех четырех случаев (исходная/уточненная модель; обучающая/тестовая выборка).

Из табл. 14 следует, что введение поправок к весам вершин МГ значительно улучшает исходную модель (по всем показателям). Таким образом, уравнение вида (14) более точно, чем уравнение (13), отражает связь между структурой и температурой кипения сульфидов.

г) Индексы удерживания алкилфенолов

В качестве исходных данных для построения модели была рассмотрена выборка алкилфенолов с их значениями индексов удерживания, N = 30, N_1 = 20, N_2 = 10 [40]. Исходные данные приведены в табл. 15.

Наилучшей моделью, отобранной на первом этапе, является уравнение вида (15):

$$y = a\sum\left(w_{ij}\right)^{-1} + b\sum\left(w_{ii}\times w_{jj}\right)^{\frac{1}{2}} + c\,. \qquad (15)$$

Далее проводится классификация вершин МГ; результаты классификации приведены в табл. 16. Атомы C, входящие в бензольное кольцо, обозначены через C^b.

На основе этой классификации были введены поправки x_1–x_9 к весам вершин. Найденные поправки x_1, x_2 , …, x_9 и коэффициенты a, b, c приведены в табл. 17.

Таким образом, наилучшая модель имеет следующий вид (16):

Таблица 14. Некоторые статистические параметры исходной/модифицированной модели для обучающей/тестовой выборки для температуры кипения сульфидов
Table 14. Some statistical parameters of the initial/modified models of the training/test samples for the boiling point of sulfides

Статистические параметры модели Statistical parameters	Исходная модель (без поправок к весам вершин МГ) Initial model (without the corrections of w_{ii} of MG)		Модифицированная модель (с поправками к весам вершин МГ) Modified model (with the corrections of w_{ii} of MG)	
	Обучающая выборка Training sample	Тестовая выборка Test sample	Обучающая выборка Training sample	Тестовая выборка Test sample
s	7.76	10.47	2.73	2.64
R	0.954	0.846	0.994	0.991
$\delta_{ср}$ / δ_{avg}	4.41%	6.77%	1.83%	1.81%
$\delta_{макс}$ / δ_{max}	12.48%	10.72%	5.60%	4.07%

Таблица 15. Алкилфенолы и их индексы удерживания [40]
Table 15. Alkylphenols and their retention indices [40]

№	Название соединения Compound	Индексы удерживания Retention indices
1	Фенол / Phenol	1281
2	2-Метилфенол / 2-Methylphenol	1354
3	3-Метилфенол* / 3-Methylphenol*	1386
4	4-Метилфенол / 4-Methylphenol	1385
5	2-Этилфенол / 2-Ethylphenol	1430
6	3-Этилфенол / 3-Ethylphenol	1483
7	4-Этилфенол* / 4-Ethylphenol*	1473
8	2,3-Диметилфенол / 2,3-Dimethylphenol	1495
9	2,4-Диметилфенол / 2,4-Dimethylphenol	1456
10	2,5-Диметилфенол / 2,5-Dimethylphenol	1453
11	2,6-Диметилфенол* / 2,6-Dimethylphenol*	1416
12	3,5-Диметилфенол* / 3,5-Dimethylphenol*	1489
13	3,4-Диметилфенол / 3,4-Dimethylphenol	1530
14	4-Изопропилфенол / 4-Isopropylphenol	1527
15	2-*н*-Пропилфенол* / 2-*n*-Propylphenol*	1502
16	3-*н*-Пропилфенол / 3-*n*-Propylphenol	1565
17	4-*н*-Пропилфенол / 4-*n*-Propylphenol	1563
18	2-Этил-4-метилфенол* / 2-Ethyl-4-methylphenol*	1523
19	2-Этил-5-метилфенол / 2-Ethyl-5-methylphenol	1529
20	2-Этил-6-метилфенол / 2-Ethyl-6-methylphenol	1485
21	3-Этил-5-метилфенол* / 3-Ethyl-5-methylphenol*	1581
22	4-Этил-2-метилфенол* / 4-Ethyl-2-methylphenol*	1539
23	4-Этил-3-метилфенол / 4-Ethyl-3-methylphenol	1608
24	2,3,4-Триметилфенол / 2,3,4-Trimethylphenol	1638
25	2,3,5-Триметилфенол / 2,3,5-Trimethylphenol	1593
26	2,3,6-Триметилфенол* / 2,3,6-Trimethylphenol*	1551
27	2,4,5-Триметилфенол* / 2,4,5-Trimethylphenol*	1593
28	3,4,5-Триметилфенол / 3,4,5-Trimethylphenol	1667
29	4-*втор*-Бутилфенол / 4-*sec*-Butylphenol	1612
30	2-*н*-Бутилфенол / 2-*n*-Butylphenol	1600

Таблица 16. Классификация вершин МГ, соответствующих молекулам алкилфенолов
Table 16. Classification of vertices of MG, corresponding to alkylphenol molecules

№	Структурный фрагмент, определяющий класс атомов Structural fragment that defines the class of atoms	Описание класса атомов Description of the atom class
1	$C^b - \mathbf{C_1^b} - C^b$	Центральный атом C_1^b связан с двумя атомами C^b The central C_1^b atom is bonded to two C^b atoms
2	$C^b - \mathbf{C_2^b}(-C)- C^b$	Центральный атом C_2^b связан с двумя атомами C^b и атомом C The central C_2^b atom is bonded to two C^b and one C atoms
3	$C^b - \mathbf{C_3}$	Центральный атом C_3 связан с атомом C^b The central C_3 atom is bonded to the C^b atom

Таблица 16. Окончание
Table 16. Continued

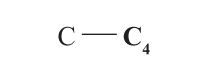
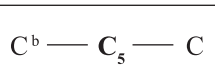
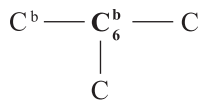
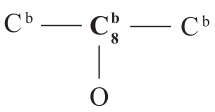
№	Структурный фрагмент, определяющий класс атомов Structural fragment that defines the class of atoms	Описание класса атомов Description of the atom class
4		Центральный атом C_4 связан с атомом C The central C_4 atom is bonded to the C atom
5		Центральный атом C_5 связан с атомом C^b и с атомом C The central C_5 atom is bonded to one C^b and one C atoms
6		Центральный атом C_6^b связан с атомом C^b и с двумя атомами C The central C_6^b atom is bonded to one C^b and two C atoms
7		Центральный атом C_7 связан с двумя атомами C The central C_7 atom is bonded to two C atoms
8		Центральный атом C_8^b связан с двумя атомами C^b и атомом O The central C_8^b atom is bonded to two C^b and one O atoms
9		Центральный атом O связан с атомом C^b The central O atom is bonded to one C^b atom

Таблица 17. Значения поправок x_1–x_9 и коэффициентов a, b, c для улучшенной модели
Table 17. Values of the corrections parameters, x_1–x_9, and the coefficients a, b, and c for the improved model

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	a	b	c
0.6198	0.5934	0.7245	0.2346	0.4256	0.5157	0.3836	0.9956	0.9777	780.9	1055.5	28.2

$$y = a\sum\left(w_{ij}\right)^{-1} + b\sum\left(\left(w_{ii}+x_n\right)\times\left(w_{jj}+x_m\right)\right)^{\frac{1}{2}} + c, \quad (16)$$

где n и m – это номера классов вершин, к которым принадлежат соответствующие вершины i и j; значения поправок x_n и x_m, а также коэффициентов a, b, c берутся из табл.17; суммирование в каждой сумме идет по всем ребрам (i,j) в МГ.

В табл. 18 приведены некоторые статистические параметры моделей для всех четырех случаев (исходная/уточненная модель; обучающая/тестовая выборка).

Из табл. 18 следует, что введение поправок к весам вершин МГ значительно улучшает первоначальную модель (по всем показателям). Таким образом, уравнение вида (16) более точно, чем уравнение (15), отражает связь между структурой и индексами удерживания алкилфенолов.

Таблица 18. Некоторые статистические параметры исходной/модифицированной модели для обучающей/тестовой выборки для индексов удерживания алкилфенолов
Table 18. Some statistical parameters of the initial/modified models of the training/test samples for the retention indices of alkylphenols

Статистические параметры модели Statistical parameters of the model	Исходная модель (без поправок к весам вершин МГ) Initial model (without the corrections of w_{ii} of MG)		Модифицированная модель (с поправками к весам вершин МГ) Modified model (with the corrections of w_{ii} of MG)	
	Обучающая выборка Training sample	Тестовая выборка Test sample	Обучающая выборка Training sample	Тестовая выборка Test sample
s	36.92	30.37	15.03	16.18
R	0.934	0.904	0.989	0.974
$\delta_{ср}$ / δ_{avg}	1.94%	2.01%	0.72%	1.00%
$\delta_{макс}$ / δ_{max}	5.86%	5.42%	2.16%	3.17%

ЗАКЛЮЧЕНИЕ

В настоящей работе разработан общий алгоритмический метод построения математических моделей связи «структура-свойство». Вышеуказанные модели строятся на основе статистического анализа данных по выборкам структур и свойств химических соединений некоторых классов и имеют вид корреляционных уравнений. Метод применим к любым классам органических соединений и любым свойствам, измеряемым количественно.

Для представления химических структур в рамках предложенного метода использованы специальные взвешенные молекулярные графы, отражающие некоторые элементы пространственного строения соответствующих молекул.

Реализация метода происходит в несколько этапов. На первом этапе предполагается, что искомое уравнение связи «структура-свойство» имеет вполне определенный вид и зависит от ряда подгоночных числовых параметров и подбираемой пары функций одной переменной. На этом этапе происходит отбор наилучшей пары функций (по определенным критериям) из некоторого множества функций или, что то же самое, наилучшей модели. При этом задаваемое множество функций можно расширять, что способствует увеличению точности наилучшей отобранной модели. Если необходимо построить более точную модель, то следует перейти ко второму этапу.

На втором этапе происходит некоторая модификация полученной наилучшей модели, направленная на улучшение ее точности. Для этой цели первоначально проводится классификация вершин молекулярного графа по химическим символам соответствующих атомов и картинам первого окружения, а также ребер, в соответствии с классами вершин, которые они соединяют. На основе полученной классификации вершин и ребер вводятся числовые «поправки» к исходным весам вершин и ребер. Конечным результатом процесса построения модели служит уравнение определенного вида с конкретными числовыми значениями всех его параметров, позволяющее рассчитать значение изучаемого свойства для любого соединения заданного класса.

Заметим, что на втором этапе можно ограничиться лишь классификацией вершин графа, а к классификации ребер переходить, только если получаемая модель, основанная на классификации вершин, не достаточно точна. При этом возможна ситуация, когда новая модель будет лучше прежней на обучающей выборке и хуже на тестовой. Еще один путь улучшить модель, получаемую на втором этапе в результате описанной выше классификации вершин МГ, это провести более детальную классификацию вершин, например, по картинам окружения второго порядка.

Кроме того, в работе проведены примеры применения разработанного метода для построения моделей связи «структура-свойство» для конкретных свойств и классов соединений, показавшие его эффективность, а также проанализирована целесообразность введения второго этапа в этом методе.

Модели связи «структура-свойство», построенные на основе вышеописанного метода, обладающие достаточно высоким качеством, могут быть использованы для расчета свойств соединений, для которых отсутствуют экспериментальные данные.

Финансовая поддержка

Данное исследование не имело какой-либо финансовой поддержки от сторонних организаций.

Financial support

This study did not have any financial support from outside organizations.

Вклад авторов

Все авторы в равной степени внесли свой вклад в исследовательскую работу.

Authors' contribution

All authors equally contributed to the research work.

Авторы заявляют об отсутствии конфликта интересов.
The authors declare no conflicts of interest.

СПИСОК ЛИТЕРАТУРЫ

1. Раевский О.А. Моделирование соотношений «структура-свойство». М.: Добросвет; 2015. 288 с. ISBN 978-5-7913-0103-11

2. Stuper A.J., Brügger W.E., Jurs P.C. Computer-Assisted Studies of Chemical Structure and Biological Function. New York: John Wiley & Sons; 1979. 220 p.

3. Розенблит А.Б., Голендер И.Е. Логико-комбинаторные методы в конструировании лекарств. Рига: Зинатне, 1983. 352 с.

4. Cherkasov A., Muratov E.N., Fourches D., Varnek A., Baskin I.I., Cronin M., Dearden J., Gramatica P., Martin Y.C., Todeschini R., Consonni V., Kuz'min V.E., Cramer R., Benigni R., Yang C., Rathman J., Terfloth L., Gastaiger J., Richard A., Tropsha A. QSAR Modeling: Where have you been? Where are you going to? *J. Med. Chem.* 2014;57(12):4977-5010. https://doi.org/10.1021/jm4004285

REFERENCES

1. Raevskii O.A. *Modelirovanie sootnoshenii "struktura–svoistvo"* (Modeling the Structure–Property Relationships). Moscow: Dobrosvet; 2015. 288 p. (in Russ.). ISBN 978-5-7913-0103-11

2. Stuper A.J., Brügger W.E., Jurs P.C. Computer-Assisted Studies of Chemical Structure and Biological Function. New York: John Wiley & Sons; 1979. 220 p.

3. Rozenblit A.B., Golender I.E. *Logiko-kombinatornye metody v konstruirovanii lekarstv* (Logico-combinatorial Methods in Drug Design). Riga: Zinatne; 1983. 352 p. (in Russ.).

4. Cherkasov A., Muratov E.N., Fourches D., Varnek A., Baskin I.I., Cronin M., Dearden J., Gramatica P., Martin Y.C., Todeschini R., Consonni V., Kuz'min V.E., Cramer R., Benigni R., Yang C., Rathman J., Terfloth L., Gastaiger J., Richard A., Tropsha A. QSAR Modeling: Where have you been? Where are you going to? *J. Med. Chem.* 2014;57(12):4977-5010. https://doi.org/10.1021/jm4004285

5. Нижний С.В., Эпштейн Н.А. Количественные соотношения «химическая структура-биологическая активность». *Успехи химии.* 1978;47(4):739-772.

6. Папулов Ю.Г., Чернова Т.И., Смоляков В.М., Поляков М.Н. Использование топологических индексов при построении корреляций «структура-свойство». *Журн. физ. химии.* 1993;67(2):203-209.

7. Филимонов Д.А., Поройков В.В., Караичева Е.И., Казарян Р.К., Будунова А.П., Михайловский Е.М., Рудницких А.В., Гончаренко Л.В., Буров Ю.В. Компьютерное прогнозирование спектра биологической активности химических соединений по их структурной формуле: система PASS. *Экспер. и клин. фармакол.* 1995;58(2):56-62.

8. Fujita T. The Application of Classical QSAR to Agrochemical Research. *International Journal of Quantitative Structure-Property Relationships (IJQSPR).* 2017;2(1):1-18. https://doi.org/10.4018/IJQSPR.2017010101

9. Torrens F., Gastellano G. QSPR Prediction of Retention Times of Methylxanthines and Cotinine by Bioplastic Evolution. *International Journal of Quantitative Structure–Property Relationships (IJQSPR).* 2018;3(1):74-87. https://doi.org/10.4018/IJQSPR.2018010104

10. Dearden J.C. The History and Development of Quantitative Structure–Activity Relationships (QSARs): Addendum. *International Journal of Quantitative Structure–Property Relationships (IJQSPR).* 2017;2(2):36-46. https://doi.org/10.4018/IJQSPR.2017070104

11. Salman M., Ahmed S., Nandy S. QSAR and Anticancer Drug Design on Benzothienopyrimidinones as Promising Pim Kinase Inhibitors Utilizing Structural Descriptors. *International Journal of Quantitative Structure–Property Relationships (IJQSPR).* 2019;4(2):82-99. https://doi.org/10.4018/IJQSPR.2019040104

12. Randić M. Comparative Regression Analysis. Regressions Based on a Single Descriptor. *Croat. Chem. Acta.* 1993;66(2):289-312.

13. Devillers J., Balaban A.T. (Eds.). Topological Indices and Related Descriptors in QSAR and QSPR. Amsterdam: Gordon and Breath Science Publishers; 1999. 811 p.

14. Todeschini R., Consonni V. Handbook of Molecular Descriptors. Weinheim: Willey-VCH; 2000. 668 p.

15. Станкевич М.И., Станкевич И.В., Зефиров Н.С. Топологические индексы в органической химии. *Успехи химии.* 1988;57(3):337-366.

16. Raevsky O.A. Molecular structure descriptors in the computer-aided design of biologically active compounds. *Russ. Chem. Rev.* 1999;68(6):505-524. https://doi.org/10.1070/RC1999v068n06ABEH000425

17. Кинг Р. Химические приложения топологии и теории графов. М.: Мир; 1987. 560 с.

18. Balaban A.T. (Ed.). Chemical Applications of Graph Theory. London: Academic Press; 1976. 389 p.

19. Trinajstić N. (Ed.). Chemical Graph Theory (2nd Edition). Boca Raton, CRC Press, 1992; 352 p.

20. Basak S.C., Magnuson V.R., Niemi G.I., Regal R.R., Veith G.D. Topological Indices: Their Nature, Mutual Relatedness and Applications. *Mathematical Modelling.* 1987;8(C):300-305. https://doi.org/10.1016/0270-0255(87)90594-X

21. Randić M. Topological Indices. In: Schleyer P.v.R., Allinger N.L., Clark T., Gasteiger J., Kollman P.A., Schaefer H.F. III, Schreiner P.R. (Eds.). The Encyclopedia of Computational Chemistry. Chichester: John Wiley & Sons; 1998. P.3018-3032.

5. Nizhnii S.V., Epshtein N.A. Quantitative "chemical structure–biological activity" relations. *Russ. Chem. Rev.* 1978;47(4):383-400. http://dx.doi.org/10.1070/RC1978v047n04ABEH002225

6. Papulov Yu.G., Chernova T.I., Smolyakov V.M., Polyakov M.N. The use of topological indices in structure–property correlations. *J. Phys. Chem.* 1993;67(2):182-188.

7. Filimonov D.A., Poroikov V.V., Karaicheva E.I., Kazaryan R.K., Budunova A.P., Mikhailovsky E.M., Rudnizkikh A.V., Goncharenko L.V., Burov Yu.V. Computer prediction of the spectrum of biological activity of chemical compounds by their structure formulae: PASS system. *Eksperimentalnaya i klinicheskaya farmakologiya = Russian Journal of Experimental and Clinical Pharmacology.* 1995; 58(2):56-62 (in Russ.).

8. Fujita T. The Application of Classical QSAR to Agrochemical Research. *International Journal of Quantitative Structure–Property Relationships (IJQSPR).* 2017;2(1):1-18. https://doi.org/10.4018/IJQSPR.2017010101

9. Torrens F., Gastellano G. QSPR Prediction of Retention Times of Methylxanthines and Cotinine by Bioplastic Evolution. *International Journal of Quantitative Structure–Property Relationships (IJQSPR).* 2018;3(1):74-87. https://doi.org/10.4018/IJQSPR.2018010104

10. Dearden J.C. The History and Development of Quantitative Structure–Activity Relationships (QSARs): Addendum. *International Journal of Quantitative Structure–Property Relationships (IJQSPR).* 2017;2(2):36-46. https://doi.org/10.4018/IJQSPR.2017070104

11. Salman M., Ahmed S., Nandy S. QSAR and Anticancer Drug Design on Benzothienopyrimidinones as Promising Pim Kinase Inhibitors Utilizing Structural Descriptors. *International Journal of Quantitative Structure–Property Relationships (IJQSPR).* 2019;4(2):82-99. https://doi.org/10.4018/IJQSPR.2019040104

12. Randić M. Comparative Regression Analysis. Regressions Based on a Single Descriptor. *Croat. Chem. Acta.* 1993;66(2):289-312.

13. Devillers J., Balaban A.T. (Eds.). Topological Indices and Related Descriptors in QSAR and QSPR. Amsterdam: Gordon and Breath Science Publishers; 1999. 811 p.

14. Todeschini R., Consonni V. Handbook of Molecular Descriptors. Weinheim: Willey-VCH; 2000. 668 p.

15. Stankevich M.I., Stankevich I.V., Zefirov N.S. Topological indices in organic chemistry. *Russ. Chem. Rev.* 1988;57(3):191-208.

16. Raevsky O.A. Molecular structure descriptors in the computer-aided design of biologically active compounds. *Russ. Chem. Rev.* 1999;68(6):505-524. https://doi.org/10.1070/RC1999v068n06ABEH000425

17. King R.B. (Ed.). Chemical Application of Topology and Graph Theory. Amsterdam: Elsevier; 1983. 494 p.

18. Balaban A.T. (Ed.). Chemical Applications of Graph Theory. London: Academic Press; 1976. 389 p.

19. Trinajstić N. (Ed.) Chemical Graph Theory (2nd Edition). Boca Raton, CRC Press, 1992; 352 p.

20. Basak S.C., Magnuson V.R., Niemi G.I., Regal R.R., Veith G.D. Topological Indices: Their Nature, Mutual Relatedness and Applications. *Mathematical Modelling.* 1987;8(C):300-305. https://doi.org/10.1016/0270-0255(87)90594-X

21. Randić M. Topological Indices. In: Schleyer P.v.R., Allinger N.L., Clark T., Gasteiger J., Kollman P.A., Schaefer H.F. III, Schreiner P.R. (Eds.). The Encyclopedia of Computational Chemistry. Chichester: John Wiley & Sons; 1998. P.3018-3032.

22. Pogliani L. How Far Are Molecular Connectivity Descriptors from IS Molecular Pseudoconnectivity Descriptors? *J. Chem. Inf. Comput. Sci.* 2001;41(3):836-847. https://doi.org/10.1021/ci000142c

23. Kier L. B. Indexes of Molecular Shape from Chemical Graphs. *Med. Res. Rev.* 1987;7:417-440.

24. Антипин И.С., Арсланов Н.А., Палюлин В.А., Коновалов А.И., Зефиров Н.С. Прогнозирование энтальпий неспецифической сольватации органических неэлектролитов. *Доклады АН.* 1993;331(2):173-176.

25. Яковенко Ю.Ю., Скворцова М.И., Михайлова Н.А. Моделирование связи между структурой и физико-химическими свойствами органических соединений на основе оптимальных атомных параметров. *Вестник МИТХТ*. 2012;7(6):110-113.

26. Randić M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* 1975;97(23):6609-6615. https://doi.org/10.1021/ja00856a001

27. Kier L.B., Hall L.H. Molecular Connectivity in Structure-Activity Analysis. N.Y.: Research Studies Press Ltd., John Wiley and Sons Inc.; 1986. 262 p.

28. Kier L.B., Hall L.H. Molecular Connectivity in Chemistry and Drug Research. N.Y.: Academic Press; 1976. 257 p.

29. Randić M., Pompe M. The Variable Connectivity Index $^{1}\chi^{f}$ Versus the Traditional Molecular Descriptors: A Comparative Study of $^{1}\chi^{f}$ Against Descriptors of CODESSA. *J. Chem. Inform. Comput. Sci.* 2001;41(3):631-638. https://doi.org/10.1021/ci000119r

30. Красных Е.Л. Прогнозирование энтальпий испарения на основе модифицированных индексов Рандича. Алканы. *Жур. структурн. химии.* 2008;49(6):1026-1033.

31. Красных Е.Л. Прогнозирование энтальпий испарения на основе модифицированных индексов Рандича. Одноатомные спирты. *Журн. структурн. химии.* 2009;50(3):557-561.

32. Красных Е.Л. Прогнозирование энтальпий испарения на основе модифицированных индексов Рандича. Многоатомные спирты. *Журн. структурн. химии.* 2009;50(4):631-637.

33. Красных Е.Л. Прогнозирование энтальпий испарения на основе модифицированных индексов Рандича. III. Карбоновые кислоты. *Журн. структурн. химии.* 2010;51(3)557-561. (2):231-236.

34. Красных Е.Л. Прогнозирование энтальпий испарения на основе модифицированных индексов Рандича. Простые эфиры. *Журн. структурн. химии*. 2012;53(2):399-403.

35. Красных Е.Л. Прогнозирование энтальпий испарения на основе модифицированных индексов Рандича. Альдегиды и кетоны. *Журн. структурн. химии.* 2013;54(4):746-750.

36. Нестерова Т.Н., Нестеров И.А. Критические температуры и давления органических соединений. Анализ состояния базы данных и развитие методов прогнозирования. Самара: Изд-во Самарский научный центр РАН; 2009. 580 с. ISBN 978-5-93424-424-9

37. Estrada E. Spectral Moments of the Edge-Adjacency Matrix of Molecular Graphs. 2. Molecules Containing Heteroatoms and QSAR Applications. *J. Chem. Inf. Comput.* Sci. 1997;37(2):320-328. https://doi.org/10.1021/ci960113v

38. Skvortsova M.I., Fedyaev K.S., Palyulin V.A., Zefirov N.S. Molecular Design of Chemical Compounds with Prescribed Properties from QSAR Models Containing the Hosoya Index. *Internet Electronic Journal of Molecular Design (IEJMD).* 2003;2(2):70-85. URL: http://www.biochempress.com/av02_0070.html

22. Pogliani L. How Far Are Molecular Connectivity Descriptors from IS Molecular Pseudoconnectivity Descriptors? *J. Chem. Inf. Comput. Sci.* 2001;41(3):836-847. https://doi.org/10.1021/ci000142c

23. Kier L. B. Indexes of Molecular Shape from Chemical Graphs. *Med. Res. Rev.* 1987;7:417-440.

24. Antipin I.S., Arslanov N.A., Palyulin V.A., Konovalov A.I., Zefirov N.S. Prediction of nonspecific solvation enthalpy for organic non-electrolites. *Doklady Akademii Nauk.* 1993;331(2):173-176 (in Russ.).

25. Yakovenko Yu.Yu., Skvortsova M.I., Mikhailova N.A. Modeling the relation between the structure and physicochemical properties of organic compounds on the basis of optimal atom parameters. *Vestnik MITHT = Fine Chemical Technologies.* 2012;7(6):110-113 (in Russ.).

26. Randić M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* 1975;97(23):6609-6615. https://doi.org/10.1021/ja00856a001

27. Kier L.B., Hall L.H. Molecular Connectivity in Structure–Activity Analysis. N.Y.: Research Studies Press Ltd., John Wiley and Sons Inc.; 1986. 262 p.

28. Kier L.B., Hall L.H. Molecular Connectivity in Chemistry and Drug Research. N.Y.: Academic Press; 1976. 257 p.

29. Randić M., Pompe M. The Variable Connectivity Index $^{1}\chi^{f}$ Versus the Traditional Molecular Descriptors: A Comparative Study of $^{1}\chi^{f}$ Against Descriptors of CODESSA. *J. Chem. Inform. Comput. Sci.* 2001;41(3):631-638. https://doi.org/10.1021/ci000119r

30. Krasnykh E.L. Prediction of vaporization enthalpy based on modified Randič indices. Alkanes. *J. Struct. Chem.* 2008;49(6):986-983. https://doi.org/10.1007/s10947-008-0170-9

31. Krasnykh E.L. Prediction of vaporization enthalpy based on modified Randič indices. I. Monohydric alcohols. *J. Struct. Chem.* 2009;50(3):556-560. https://doi.org/10.1007/s10947-009-0084-1

32. Krasnykh E.L. Prediction of vaporization enthalpy based on modified Randič indices. II. Polyatomic alcohols. *J. Struct. Chem.* 2009;50(4):599-605. https://doi.org/10.1007/s10947-009-0094-z

33. Krasnykh E.L. Prediction of vaporization enthalpy based on modified Randič indices. III. Carbonic acides. *J. Struct. Chem.* 2010;51(2):217-222. https://doi.org/10.1007/s10947-010-0034-y

34. Krasnykh E.L. Prediction of vaporization enthalpy based on modified Randič indices. Ethers. *J. Struct. Chem.* 2012;53(2):383-387. https://doi.org/10.1134/S0022476612020266

35. Krasnykh E.L. Prediction of vaporization enthalpy based on modified Randič indices. Aldehydes and ketones. *J. Struct. Chem.* 2013;54(4):792-796. https://doi.org/10.1134/S0022476613040203

36. Nesterova T.N., Nesterov I.A. *Kriticheskie temperatury i davleniya organicheskikh soedinenii. Analiz sostoyaniya bazy dannykh i razvitie metodov prognozirovaniya* (Critical Temperatures and Pressures of Organic Compounds. The Analysis of Data Base and Development of Prediction Methods). Samara: Samarskii nauchnyi tsentr RAN Publ.; 2009. 580 p. (in Russ.). ISBN 978-5-93424-424-9

37. Estrada E. Spectral Moments of the Edge-Adjacency Matrix of Molecular Graphs. 2. Molecules Containing Heteroatoms and QSAR Applications. *J. Chem. Inf. Comput. Sci.* 1997;37(2):320-328. https://doi.org/10.1021/ci960113v

39. Zefirov N.S., Palyulin V.A. QSAR for Boiling Points of "Small" Sulfides. Are the "High-Quality Structure-Property-Activity Regressions" the Real High Quality QSAR Models? *J. Chem. Inf. Comput. Sci.* 2001;41(4):1022-1027. https://doi.org/10.1021/ci0001637

40. Zefirov N.S., Palyulin V.A. Fragmental Approach in QSPR. *J. Chem. Inf. Comput. Sci.* 2002;42(5):1112-1122. https://doi.org/10.1021/ci020010e

38. Skvortsova M.I., Fedyaev K.S., Palyulin V.A., Zefirov N.S. Molecular Design of Chemical Compounds with Prescribed Properties from QSAR Models Containing the Hosoya Index. *Internet Electronic Journal of Molecular Design (IEJMD).* 2003;2(2):70-85. URL: http://www.biochempress.com/av02_0070.html

39. Zefirov N.S., Palyulin V.A. QSAR for Boiling Points of "Small" Sulfides. Are the "High-Quality Structure-Property-Activity Regressions" the Real High Quality QSAR Models? *J. Chem. Inf. Comput. Sci.* 2001;41(4):1022-1027. https://doi.org/10.1021/ci0001637

40. Zefirov N.S., Palyulin V.A. Fragmental Approach in QSPR. *J. Chem. Inf. Comput. Sci.* 2002;42(5):1112-1122. https://doi.org/10.1021/ci020010e

Об авторах:

Шулаева Надежда Александровна, студентка кафедры высшей и прикладной математики Института тонких химических технологий им. М.В. Ломоносова ФГБОУ ВО «МИРЭА – Российский технологический университет» (119571, Россия, Москва, пр. Вернадского, д. 86). E-mail: akacija@inbox.lv. https://orcid.org/0000-0002-5974-5213

Скворцова Мария Ивановна, доктор физико-математических наук, доцент, заведующий кафедрой высшей и прикладной математики Института тонких химических технологий им. М.В. Ломоносова ФГБОУ ВО «МИРЭА – Российский технологический университет» (119571, Россия, Москва, пр. Вернадского, д. 86). E-mail: skvorivan@mail.ru. Scopus Author ID 6603801652, https://orcid.org/0000-0002-6179-8875

Михайлова Наталия Александровна, старший преподаватель кафедры высшей и прикладной математики Института тонких химических технологий им. М.В. Ломоносова ФГБОУ ВО «МИРЭА – Российский технологический университет» (119571, Россия, Москва, пр. Вернадского, д. 86). E-mail: essen.05@mail.ru. https://orcid.org/0000-0002-8888-0190

About the authors:

Nadezhda A. Shulaeva, Student, Department of Higher and Applied Mathematics, M.V. Lomonosov Institute of Fine Chemical Technologies, MIREA – Russian Technological University (86, Vernadskogo pr., Moscow, 119571, Russia). E-mail: akacija@inbox.lv. https://orcid.org/0000-0002-5974-5213

Mariya I. Skvortsova, Dr. of Sci. (Physics and Mathematics), Associate Professor, Head of the Department of Higher and Applied Mathematics, M.V. Lomonosov Institute of Fine Chemical Technologies, MIREA – Russian Technological University (86, Vernadskogo pr., Moscow, 119571, Russia). E-mail: skvorivan@mail.ru. Scopus Author ID 6603801652, https://orcid.org/0000-0002-6179-8875

Nataliya A. Mikhailova, Senior Lecturer, Department of Higher and Applied Mathematics, M.V. Lomonosov Institute of Fine Chemical Technologies, MIREA – Russian Technological University (86, Vernadskogo pr., Moscow, 119571, Russia). E-mail: essen.05@mail.ru, https://orcid.org/0000-0002-8888-0190

Поступила: 10.01.2020; получена после доработки: 24.02.2020; принята к опубликованию: 16.10.2020.
The article was submitted: January 10, 2020; approved after reviewing: February 24, 2020; accepted for publication: October 16, 2020.