УДК 544.354.081.7:004.021.

Оценка температуры плавления низкомолекулярных органических соединений на основе анализа количественной взаимосвязи «структура–свойство»

***Мансурова М.Г.**, студент*

кафедра «Теоретическая информатика и компьютерные технологии»

Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана

Научный руководитель: Дубанов А.В.,

Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана

qrcs@mail.ru

Введение. Прогнозирование температуры плавления является сложной задачей из области исследований количественных соотношений «структура-свойство» [5, 23]. Известно, что результаты прогнозирования напрямую связаны с качеством обучающей выборки; также большое значение имеют знания о точном представлении химической структуры веществ [5]. Однако доступный набор данных для предсказания температуры плавления ограничен примерно 50000 соединениями [5], в результате чего возникают сложности при прогнозировании.

Актуальность таких исследований состоит в том, что их результаты позволяют предсказывать свойства новых соединений, появляющихся в большом количестве при создании новых материалов. Стоит также отметить, что температура плавления имеет отношение к прогнозированию токсичности веществ, а также связана с некоторыми физическими свойствами химических соединений, например, температурой кипения или давлением пара [5].

Несмотря на то, что задача прогнозирования температуры плавления считается хорошо проработанной, в настоящий момент не существует удовлетворительного решения данной проблемы [7]. Для веществ, схожих по своей структуре, наилучшим результатом являлся

прогноз с ошибкой в 20 °C [3, 4], для остальных случаев – прогноз с ошибкой в 30 °C [7, 11, 9, 20]. Здесь и далее по тексту под ошибкой подразумевается стандартное отклонение.

Цель данной работы заключается в поиске такого набора молекулярных дескрипторов, при использовании которого QSPR-моделирование позволило бы найти наилучшую модель для прогноза температуры плавления.

Объекты и методы. В качестве данных для обучающей и контрольной выборки были взяты вещества из базы данных Alfa Aesar Curated (8738 веществ) [14]. Она является одной из самых распространенных среди баз данных, содержащих большое количество веществ с известной температурой плавления, установленной экспериментально, и находящихся в открытом доступе [14]. В ней есть также информация о названиях соединений, их SMILES, CSID, NSN, номерах CAS и данных об источниках.

Вычисления 2D-дескрипторов были выполнены с помощью библиотеки RDKit для языка Python в среде исполнения PyCharm [16] на ПК с параметрами Intel Core i7, 2,40 ГГц, 5006 Мб RAM под управлением операционной системы Ubuntu Linux (64 bit). Для вычисления 3D-дескрипторов модели 3D структур исследуемых соединений были получены с помощью библиотеки Open Babel и языка Python на ПК с операционной системой openSUSE Linux Leap 42.1.

Работа с QSPR-моделями и их оптимизация была выполнена средствами языка R [13, 21]. Также был использован электронный ресурс OPEN Notebook Science для прогнозирования температуры плавления некоторых химических веществ [10]. Прогнозирование температуры с помощью этого ПО было выполнено для соединений из базы данных DrugBank [12].

В табл. 1 представлены параметры, характеризующие рассматриваемую выборку Alfa Aesar Curated: минимум, максимум, нижний Q1, средний Q2 (медиана) и верхний Q3 квартили. Для построения моделей были использованы следующие выборки соединений с известными температурами плавления: вся Alfa Aesar Curated (8738 вещ-в) – выборка A. Выборкой B назовем группу первых 5826 веществ из A (первые 2/3 веществ), выборкой C – группу последних 2912 веществ из A (последняя 1/3 веществ).

Таблица 1

Характеристика выборки Alfa Aesar Curated

Выборка	Min, °C	Max, °C	Q1, °C	Q2 (медиана), °C	Q3, °C
A	-170,0	417,0	38,0	83,0	141,5
B	-157,0	379,0	43,5	86,0	144,0
C	-170,0	417,0	19,0	77,0	135,5

Поиск наилучших двумерных дескрипторов для QSPR-моделирования. Библиотека RDKit языка Python предоставляет следующий набор молекулярных дескрипторов: фрагментные в количестве 85 дескрипторов (fr_Al_COO, fr_Al_OH, fr_Al_OH_noTert и др.), физико-химические (MolLogP, MolMR, MolWt, PEOE_VSA1, SMR_VSA1, SlogP_VSA1, EState_VSA1, VSA_EState1, TPSA и др.), квантово-химические (NumValenceElectrons) и другие (HeavyAtomCount, NHOHCount, NOCount, NumHAcceptors, NumHDonors, NumHeteroatoms, NumRotatableBonds, RingCount). Названия дескрипторов приведены согласно документации к пакету RDKit [16].

Все дескрипторы были разбиты на группы из коррелирующих между собой дескрипторов. Затем в каждой группе был выбран наиболее значимый для данной выборки [1]. Из таких дескрипторов был собран первый набор молекулярных дескрипторов: NumRadicalElectrons, MinPartialCharge, MinEStateIndex, MinAbsEStateIndex, BalabanJ, HallKierAlpha, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, PEOE_VSA13, PEOE_VSA2, PEOE_VSA3, PEOE_VSA5, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, SMR_VSA2, SMR_VSA4, SMR_VSA8, SlogP_VSA1, SlogP_VSA10, SlogP_VSA11, SlogP_VSA3, SlogP_VSA4, SlogP_VSA9, EState_VSA11, EState_VSA2, EState_VSA3, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA7, EState_VSA8, VSA_EState1, VSA_EState10, VSA_EState2, VSA_EState3, VSA_EState4, VSA_EState5, VSA_EState6, VSA_EState7, VSA_EState8, FractionCSP3, NumAromaticHeterocycles, NumHDonors, NumHeteroatoms, NumRotatableBonds, NumSaturatedRings.

Полученный набор из 48 дескрипторов позволил построить (средствами языка R) регрессионную модель с коэффициентом детерминации R^2 = 0,83 и $p << 0,05$. В качестве

обучающей выборки для построения модели была взята выборка B (5826 соединений), для контрольной – выборка C (2912 веществ). Значение R^2 для зависимости между экспериментальными значениями температуры плавления веществ контрольной выборки и предсказанными, вычисленными с помощью модели, оказалось равным 0,54.

При использовании электронного ресурса OPEN Notebook Science [10] для прогнозирования температуры плавления низкомолекулярных органических соединений результаты оказались хуже (R^2 = 0,45 с ошибкой в 22 °C). Химические вещества были взяты из базы данных DrugBank [12]. В качестве данных для вычислений были выбраны преимущественно низкомолекулярные органические соединения, используемые в качестве лекарственных средств. Проведение обучения на подобных веществах позволяет строить модели для прогнозирования лекарственных свойств для групп новых синтезируемых соединений [18]. Также, температура плавления, в том числе и предсказанная, может выступать в роли дескриптора при прогнозе ряда важных свойств биологически активных веществ, например, ADME-свойств (абсорбция, распространение в кровеносной системе, метаболизм, выведение из организма) [2, 5].

На рис. 1 представлены графики зависимости между экспериментальными и предсказанными значениями температуры плавления для этих двух исследований.

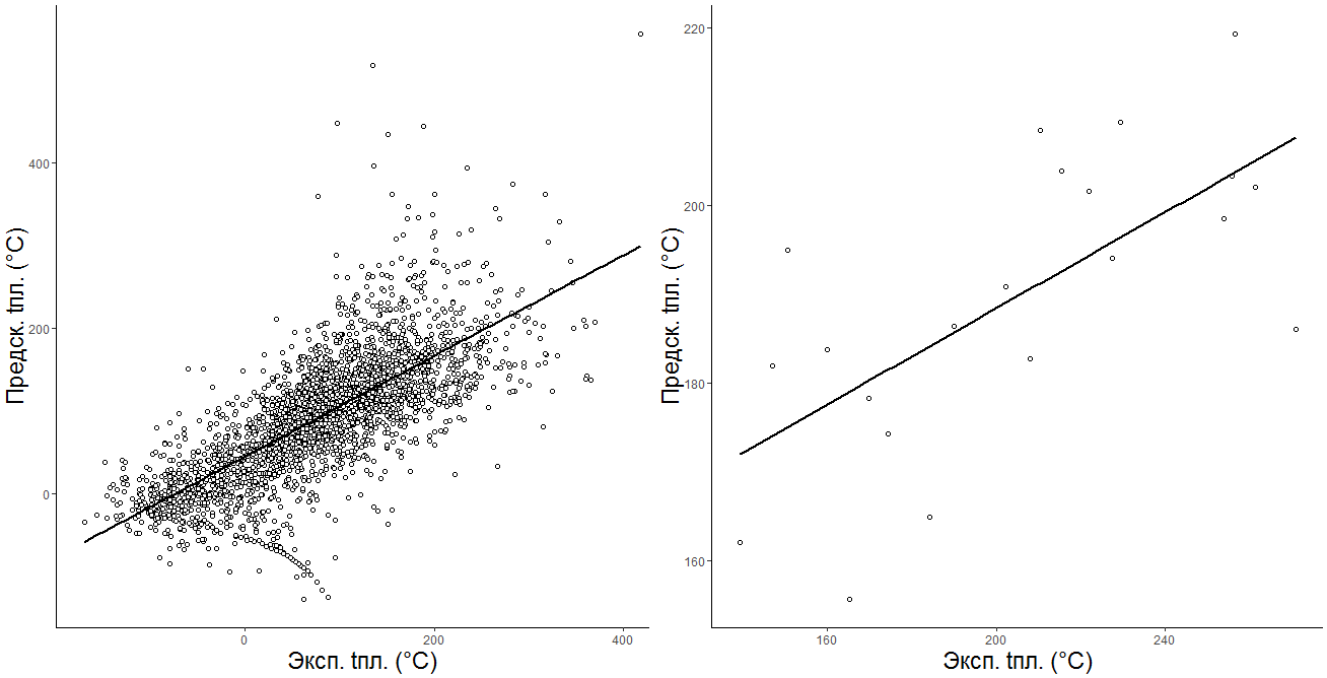

Рис. 1. *График слева – график зависимости предсказанной температуры плавления от фактической, первый набор дескрипторов. Уравнение прямой: y = 0,6x + 45,1; коэффициент детерминации:* r^2*=0,54. График справа – график зависимости предсказанной температуры плавления от фактической, вычисленной с помощью электронных ресурсов [10]. Уравнение прямой: y = 0,3x + 134; коэффициент детерминации:* r^2*=0,45.*

Таким образом можно сделать вывод, что прогноз по выявленному набору дескрипторов (далее – набор дескрипторов №1) можно считать приемлемым для прогнозирования температуры плавления низкомолекулярных соединений, так как R^2 для такого моделирования оказался лучше R^2, характеризующего выборку, полученную с помощью существующих электронных ресурсов для прогнозирования температуры плавления.

Поиск значимых дескрипторов. Для поиска дескрипторов, оказывающих большое влияние на поведение QSPR-модели для исходной выборки, было проведено обучение модели на наборе из всех дескрипторов библиотеки RDKit языка Python. Дескрипторы были выбраны по принципу наименьшего значения величины, обратной степени значимости гипотезы, и наибольшего значения достоверности коэффициента регрессии. Полученный набор дескрипторов (далее – набор дескрипторов №2) оказался следующим: BalabanJ, BertzCT, Chi0n, Chi0v, Chi1, Chi3n, HallKierAlpha, Kappa2, LabuteASA, PEOE_VSA1, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, PEOE_VSA13, PEOE_VSA14, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, SMR_VSA1, SMR_VSA10, SMR_VSA3, SMR_VSA7, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, EState_VSA10, VSA_EState8, VSA_EState9, HeavyAtomCount, NumHDonors, NumHeteroatoms, NumRotatableBonds, fr_Al_OH, fr_Ar_OH, fr_NH0, fr_NH1, fr_NH2, fr_aldehyde, fr_alkyl_carbamate, fr_alkyl_halide, fr_amide, fr_aryl_methyl, fr_azide, fr_halogen, fr_hdrzone, fr_imidazole, fr_lactone, fr_nitro_arom, fr_para_hydroxylation, fr_phenol, fr_phos_acid, fr_piperdine, fr_sulfone, fr_thiazole, fr_thiophene, fr_unbrch_alkane.

При использовании такого набора дескрипторов на выборках B и C в качестве обучающей и контрольной соответственно была построена регрессионная модель с коэффициентом детерминации, равным 0,88. Затем значимые дескрипторы набора №2 были добавлены к первому набору №1 дескрипторов. Таким образом был найден набор дескрипторов

№3 и также применен к тем же выборкам B и C. В результате была построена модель с коэффициентом детерминации, равным 0,89.

Значения R^2 для результатов прогнозирования оказались следующими: 0,67 (для набора №2 дескрипторов) и 0,69 (для набора дескрипторов №3). Графики зависимостей для данных исследований представлены на рис. 2.

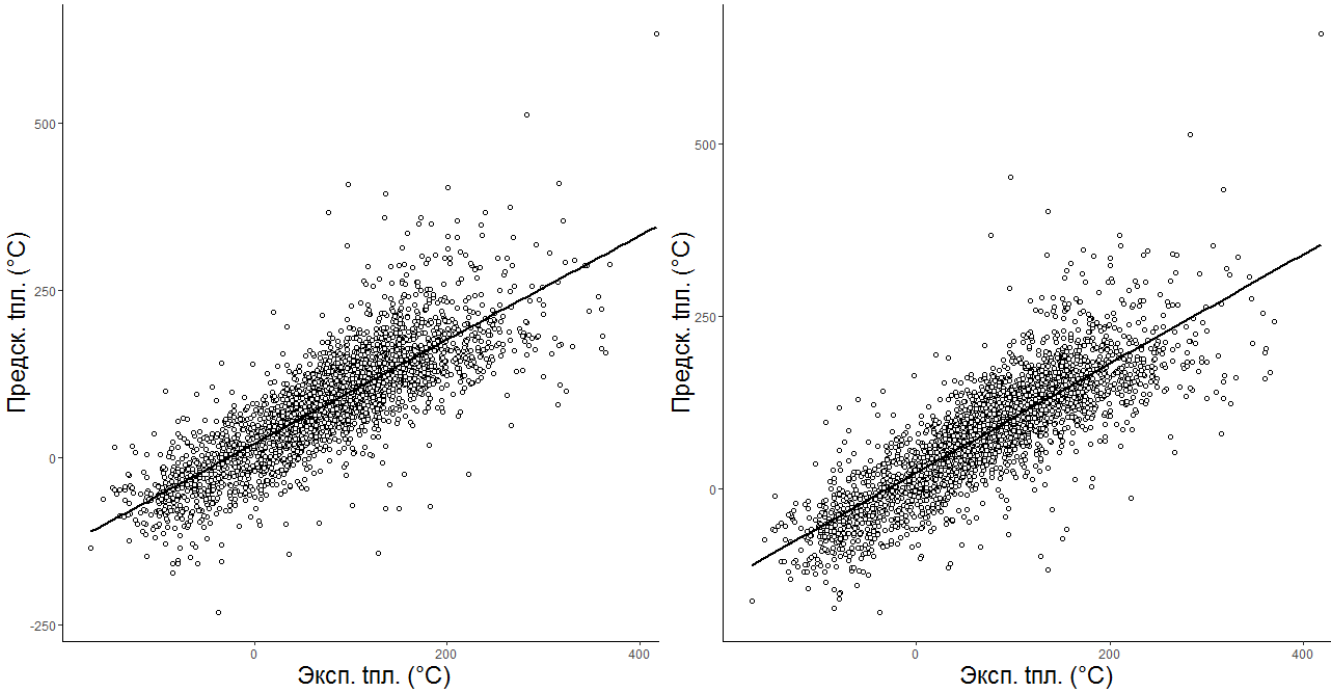

Рис. 2. *График слева – график зависимости предсказанной температуры плавления от фактической, второй набор дескрипторов. Уравнение прямой: y = 0,78x + 20,8; коэффициент детерминации: r^2=0,67. График справа – график зависимости предсказанной температуры плавления от фактической, третий набор дескрипторов. Уравнение прямой: y = 0,79x + 23,7; коэффициент детерминации: r^2=0,69.*

Данные графики позволяют заметить, что при использовании третьего набора дескрипторов для прогнозирования – значения температуры плавления оказываются меньше, чем значения температуры, полученной при использовании набора дескрипторов №2. Также было найдено, что значения R^2 оказались лучше, чем при построении зависимостей на первом наборе дескрипторов, а значит, наборы дескрипторов №2 и №3 более приемлемы для прогноза, чем набор дескрипторов №1.

Физико-химические свойства веществ как молекулярные дескрипторы. В качестве двумерных дескрипторов были рассмотрены предсказанные значения некоторых свойств веществ из базы данных DrugBank [12]: липофильность (logP), растворимость в воде (logS), топологическая полярная площадь поверхности (TPSA), способность к преломлению (Refractivity), поляризуемость (Polarizability). Назовем этот набор дескрипторов набором №4. Представленные свойства примечательны тем, что в рассматриваемой базе данных они известны для многих веществ (таким образом выборка соединений с рассматриваемыми свойствами оказалась наибольшей). Прогноз представленных величин включает в себя расчет ряда структурных и топологических дескрипторов.

Стоит отметить, что коэффициент корреляции между температурой плавления и площадью полярной поверхности молекулы равен 0,4. При попытке построить модель зависимости между температурами плавления веществ и их значениями TPSA была получена модель с коэффициентом детерминации, равным 0,63. Значение R^2 оказалось равным 0,05 – что не является хорошим результатом. В дальнейшем значение TPSA для построения модели было использовано совместно с другими дескрипторами.

При использовании таких дескрипторов на выборках B и C в качестве обучающей и контрольной соответственно была получена модель с коэффициентом детерминации, равным 0,72. График зависимости для этой модели представлен на рис. 3.

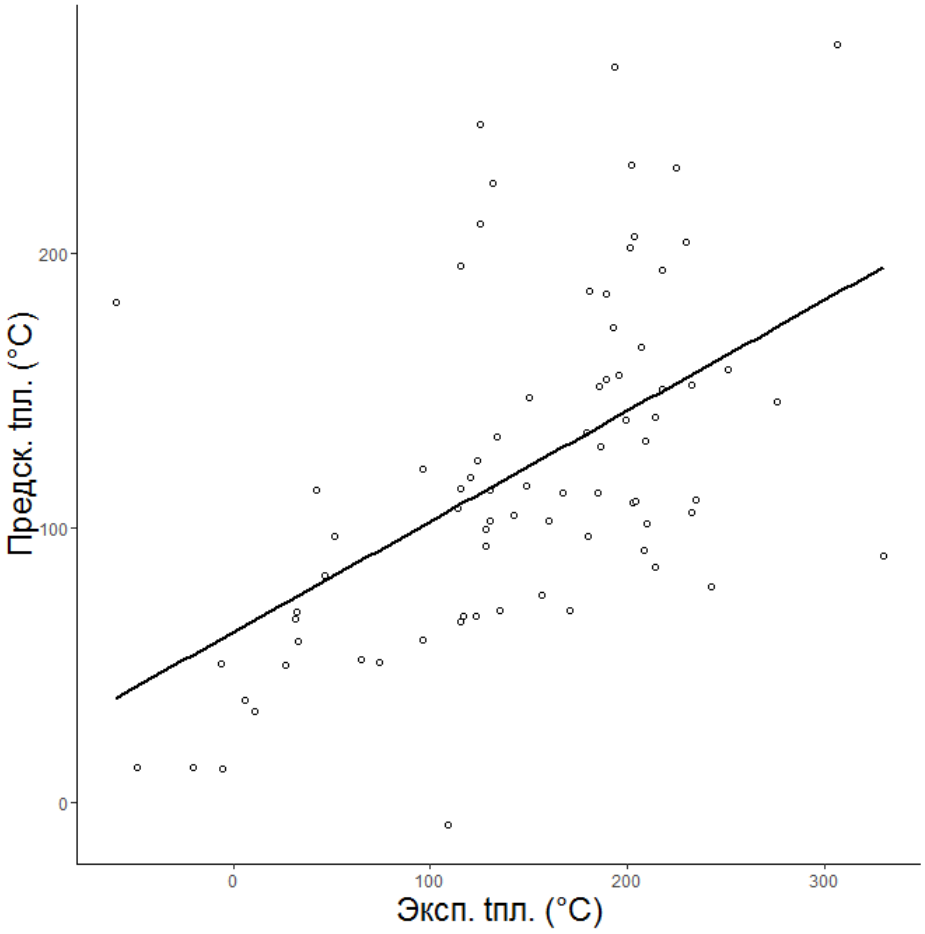

Рис. 3. *График зависимости предсказанной температуры плавления от фактической, четвертый набор дескрипторов (из предсказанных свойств веществ). Уравнение прямой: y = 0,4x + 62; коэффициент детерминации: r^2=0,29.*

Значение R^2 для выборки оказалось небольшим (всего 0,29), а значит выбранный набор дескрипторов для практического применения не пригоден.

Таким образом было найдено, что наборы дескрипторов №2 и №3 являются наилучшими для прогнозирования температуры плавления низкомолекулярных органических соединений.

Применение двумерных дескрипторов к выборке, разбитой на группы по элементному составу.

Исходная выборка (Alfa Aesar Curated) была разбита на группы веществ, близких по атомарному составу:

– вещества, содержащие только атомы углерода и водорода (1 группа, 29 веществ);

– вещества, содержащие только атомы углерода, кислорода и водорода (2 группа, 81 вещество);

– вещества, содержащие только атомы углерода, серы, кислорода и водорода (3 группа, 117 веществ);

– вещества, содержащие только атомы углерода, азота и водорода (4 группа, 55 веществ);

– вещества, содержащие только атомы углерода, азота, кислорода и водорода (5 группа, 119 веществ);

QSPR-модели для этих веществ были построены с использованием первого, второго и третьего наборов дескрипторов; в качестве обучающей выборки – первые 2/3 веществ из группы, в качестве контрольной – последняя 1/3 веществ группы. Значения R^2 и стандартных отклонений σ для моделей представлены в табл. 2 и 3 соответственно.

Таблица 2

Значения R^2 для моделей

Номер группы	Набор №1	Набор №2	Набор №3
1	0,997	0,996	0,996
2	0,893	0,969	0,923
3	0,813	0,666	0,880
4	0,636	0,590	0,531
5	0,661	0,034	0,079

Таблица 3

Стандартные отклонения (σ) для моделей

Номер группы	Набор №1	Набор №2	Набор №3
1	3,41	4,06	4,06
2	20,28	9,84	16,00
3	19,71	48,28	12,33

4	42,72	22,94	36,89
5	26,98	3937,40	18058,70

Представленные таблицы позволяют увидеть, какие наборы дескрипторов являются наиболее пригодными для каждой из групп выборок.

Наихудшие результаты были получены для пятой группы веществ (содержащих углерод, азот, кислород и водород). На наш взгляд, такое снижение точности прогноза обусловлено следующим. Соединения, содержащие азот и кислород, часто являются полярными. Эти атомы часто входят в состав функциональных групп, способных к образованию донорно-акцепторных и ионных связей. Образование таких связей является фактором, повышающим температуру плавления вещества. Отсюда, можно предположить, что используемый набор дескрипторов в недостаточной степени учитывает этот фактор.

График зависимости рассматриваемой модели представлен на рис. 4.

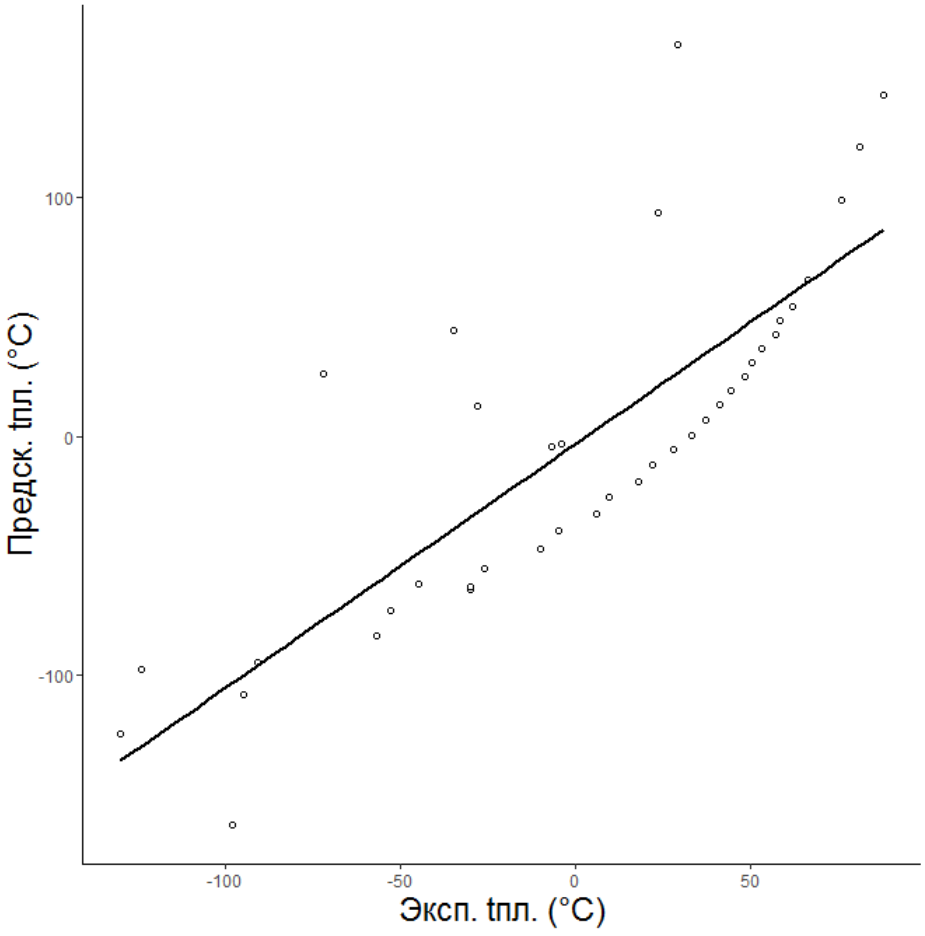

Рис. 4. *График зависимости предсказанной температуры плавления от фактической для пятой группы веществ, первый набор дескрипторов. Уравнение прямой: y = 1,02x - 3,07; коэффициент детерминации:* r^2*=0,66.*

Трехмерные дескрипторы. Особенность геометрических трехмерных молекулярных дескрипторов состоит в том, что они предоставляют информацию о трехмерной структуре молекулы, а также о ее ориентации в пространстве. Такие дескрипторы часто используются для прогноза биологической активности соединений методами QSPR [8].

Зависимость между температурой плавления вещества, характером кристаллической решетки и межмолекулярными взаимодействиями на данный момент изучена плохо, однако предполагается, что в некоторых случаях эти параметры взаимосвязаны (например, для гомологических рядов алифатических соединений) [23]. Обосновать предположение о зависимости между температурой плавления и кристаллической решеткой соединения можно тем, что кристаллическая решетка определяет близость расположения молекул в рассматриваемом веществе и тип междумолекулярной связи – что влияет на температуру.

В качестве трехмерных дескрипторов были использованы следующие значения: межатомные расстояния в молекуле (максимум, минимум, среднее расстояние между молекулами, медиана, среднеквадратическое отклонение (характеризует симметрию 3D-структуры молекулы), расстояния от геометрического центра молекулы (максимум, минимум, среднее, медиана, стандартное отклонение), максимальный и минимальный парциальные заряды, параметры молекулы, связанные с ее объемом (молярная масса, объем бокса – параллелепипеда, в котором может быть размещена трехмерная структура молекулы, молярное отношение массы к объему этого бокса, объем молекулы в $Å^3$, отношение объема молекулы к объему бокса), центр масс молекулы (наименьшая и наибольшая атомная масса, расстояние от центра масс до геометрического центра молекулы), параметры молекулы как диполя (модуль электрического дипольного момента, расстояние от центра молекулы до атома с наибольшим положительным зарядом (без учета заряда (dip_pos_l) / с учетом заряда), расстояние от центра молекулы до атома с наибольшим (по модулю) отрицательным зарядом (без учета заряда (dip_neg_l) / с учетом заряда), угол между dip_pos_l и dip_neg_l). На наш взгляд, эти величины характеризуют следующее: межатомные расстояния и парциальные заряды могут отвечать за асимметрию молекулы, дескрипторы, связанные с объемом, отражающие расстояния от

геометрического центра молекулы или определяющие центр масс, дают оценку – насколько плотно отдельные молекулы «упакованы» в твердом состоянии. Параметры молекулы как диполя могут отражать возможность электростатических межмолекулярных взаимодействий и асимметрию молекулы.

Модель была применена к веществам исходной выборки (Alfa Aesar Curated) в размере 7315 веществ (качестве обучающей и контрольной – все соединения выборки) и при использовании таких дескрипторов получилась следующей: с коэффициентом детерминации, равным 0,38, и R^2, равным 0,38.

Результаты прогнозирования представлены на рис. 5.

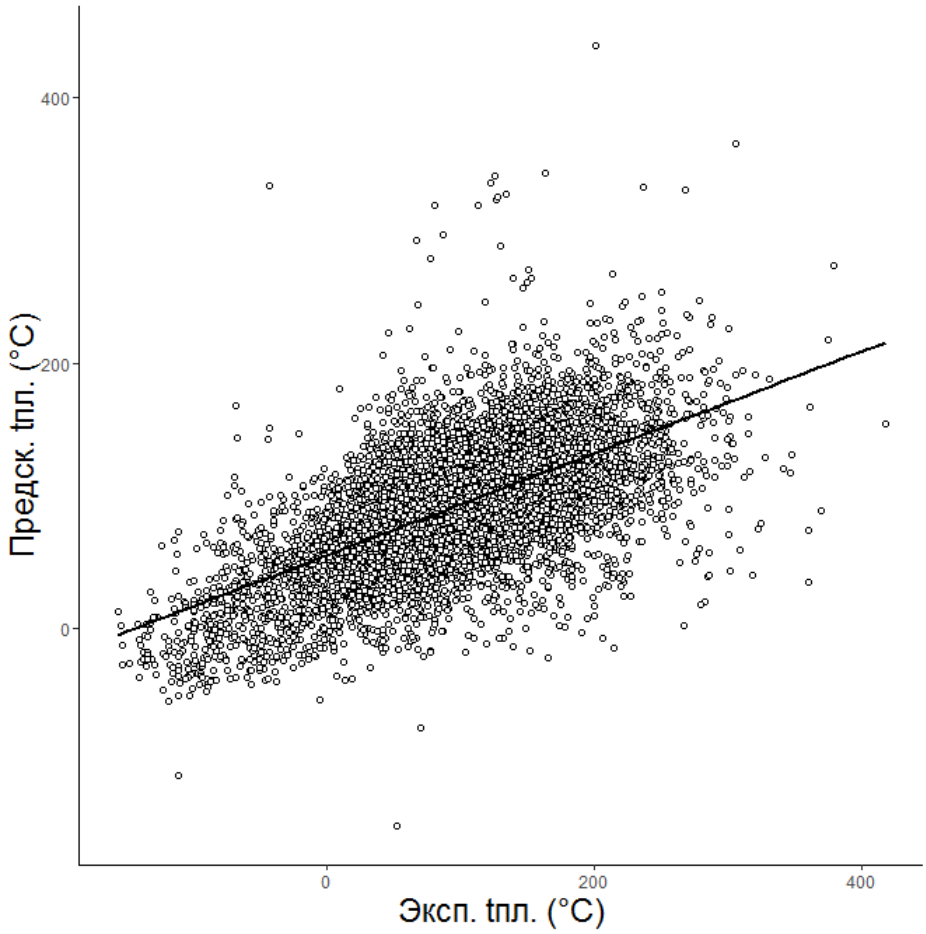

Рис. 5. *График зависимости предсказанной температуры плавления от фактической для 7315 веществ Alfa Aesar Curated, трехмерные дескрипторы. Уравнение прямой: y = 0,38x + 55,2; коэффициент детерминации: r^2=0,38.*

Значение R^2 (0,38) оказалось хуже, чем значения R^2, полученные при использовании наборов дескрипторов №1, №2, №3 (R^2=0,54, R^2=0,67 и R^2=0,69 соответственно) и немного лучше, чем при использовании набора №4 (R^2=0,29).

Следует отметить, что при использовании разных веществ для обучающей и контрольной выборок значение R^2 будет, предположительно, хуже. Таким образом было найдено, что выбранный набор трехмерных дескрипторов плохо пригоден для практического применения.

Характеристики полученных выборок. В табл. 4 представлены характеристики выборок (минимум, максимум, стандартная ошибка, стандартное отклонение, квартили (нижний Q1, средний Q2 (медиана) и верхний Q3)), состоящих из предсказанных значений температуры плавления и полученных с помощью всех рассмотренных наборов дескрипторов. В табл. 5 представлены значения R^2 для каждой из выборок. На обоих таблицах присутствует выборка, полученная с помощью электронного ресурса OPEN Notebook Science (OPEN N. S.).

Таблица 4

Характеристики полученных выборок

Набор дескрипторов	Min, °C	Max, °C	Стандартная ошибка, °C	σ, °C	Q1, °C	Q2 (медиана), °C	Q3, °C
№1	-127,24	555,46	50,60	39,95	41,57	93,39	135,95
№2	-231,78	633,33	49,31	35,36	26,04	80,51	134,29
№3	-178,62	659,26	48,54	35,03	27,69	87,75	139,38
№4	-8,27	276,15	52,11	48,51	76,09	112,57	153,37
OPEN N. S.	-155,70	219,30	12,59	21,90	182,00	190,90	202,00
3D-дескр-ры	-148,77	439,50	38,78	39,56	58,64	89,14	120,69

Таблица 5

Значения R^2 для выборок

Набор дескрипторов	Значение R^2
№1	0,54

№2	0,67
№3	0,69
№4	0,29
OPEN N. S.	0,45
3D-дескр-ры	0,38

Таким образом, исходя из значений стандартных ошибок и значений R^2 можно видеть, что наилучшие результаты были достигнуты на наборах дескрипторов №2 и №3.

Другие методы. В качестве других способов для прогнозирования температуры плавления можно рассмотреть следующие варианты:

– Исследовать зависимость между температурой плавления соединений и их конформационных свойств [17] (возможно – совместно с трехмерными дескрипторами);

– Использовать метод «ближайшего соседа» для нахождения зависимости между структурой вещества и его температурой плавления. В качестве меры сходства можно использовать меру Танимото-Жаккара [19, 15];

– Рассмотреть интерполяционные методы [6]: метод IDW, интерполяция по Шепарду [6], в качестве меры сходства использовать меру Танимото-Жаккара или расстояния в пространстве дескрипторов RDKit;

– Использовать процедуру «скользящего контроля» при построении моделей и прогнозировании [22];

Результаты, полученные при использовании данных методов, представлены в табл. 6. Исследования были проведены на выборке Alfa Aesar Curated. В качестве значений *x* и *y* для уравнений задаются фактические и предсказанные значения температуры плавления соответственно. Значение меры Танимото-Жаккара будем обозначать символом «T». Метод №1 – метод ближайшего соседа, №2 - метод ближайшего соседа с мерой T <= 0,95 (прогноз удалось выполнить для 1040 соединений из 8739, иначе говоря, только для 1/8 всех соединений имелся близкий сосед). Метод №3 подразумевает под собой интерполяцию по Шепарду, в качестве меры сходства взята мера T. Метод №4 – интерполяция по Шепарду, в качестве меры сходства – расстояния в пространстве дескрипторов RDKit (выполняем интерполяцию по 10 ближайшим соединениям; таких соединений в выборке оказалось 99,5%).

Таблица 6

Результаты, полученные при использовании других методов

Метод	Уравнение	Стандартная ошибка, °C	Значение R^2	Значение *p*
№1	y = 0,65x + 22,41	64,21	0,40	<< 0,05
№2	y = 0,71x + 20,83	56,50	0,50	<< 0,05
№3	y = 0,59x + 29,28	44,3	0,54	<< 0,05
№4	y = 0,51x + 43,30	43,25	0,48	<< 0,05

Таким образом, было найдено, что метод ближайшего соседа для практического применения не пригоден, так как ошибка очень велика. Лучшие результаты получились при использовании методов под номерами №3 и №4.

Заключение. В ходе работы было показано, что прогнозирование температуры плавления с помощью моделей, не ограниченных группой или классом соединений, не дает удовлетворительных результатов (OPEN Notebook Science, значение R^2=0,45, значение ошибки (стандартного отклонения) = 21,90) – слишком малое значение R^2 (меньше 0,5).

Было проведено прогнозирование для веществ, состоящих из классов органических соединений, больших чем гомологические ряды, но близких по своей структуре. Полученные значения ошибок находились в интервале 3,4-27,0 °C при значениях R^2 в интервале *0,636-0,997*. Такие значения можно считать допустимыми, так как наилучшим результатом для групп веществ, близких по атомарному составу, считается прогноз с ошибкой (стандартным отклонением) в 20 °C [3, 4].

Прогнозирование для веществ всей выборки также оказалось успешным: ошибки составили 35 °C при значениях R^2, равных 0,67 и 0,69 (на наборах №2 и №3 дескрипторов соответственно) [7, 11, 9, 20].

Наилучший результат с ошибкой в 3,4 °C был достигнут для выборки веществ, состоящих из углерода и водорода на наборе дескрипторов №1. Таким образом найденный набор дескрипторов является наиболее пригодным для практического применения при условии наличия выборок, состоящих только из углерода и водорода. Можно заметить, что ионные и донорно-акцепторные связи между молекулами таких веществ не образуются. Температура

плавления в этом случае определяется преимущественно массой и геометрией молекулы, что находит отражение на наборе дескрипторов.

Прогнозирование с использованием предположительного набора 3D-дескрипторов оказалось неудачным ($R^2=0,38$). Однако, трехмерные дескрипторы можно использовать совместно с другими типами дескрипторов (например, с дескрипторами, описывающими конформационные свойства соединений) и, предположительно, получать более точные модели для прогноза. Таким образом, от использования трехмерных дескрипторов отказываться не стоит.

В результате были найдены наилучшие для использования в прогнозировании температуры плавления наборы дескрипторов, применимые к выборкам, не ограниченным одним классом соединений – наборы дескрипторов №2 и №3 библиотеки RDKit.

В свою очередь, мы можем рекомендовать использование пакета RDKit для прогноза температуры плавления органических соединений методами количественного анализа взаимосвязи «структура-свойство».

ЛИТЕРАТУРА

1. Boris Johnson-Restero, Leonardo Pacheco-Londono, Jesus Olivero-Verbel, Molecular Parameters Responsible for the Milting Point of 1,2,3-Diazaborine Compounds. – J. Chem. Inf. Comput. Sci.. – 2003 – №43(5) – pp1513-1519
2. Edward H. Kerns, Li Di, Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization. – Academic Press. – 2008 – 552p
3. Guijie Liang Jie Xu Li Liu, QSPR analysis for melting point of fatty acids using genetic algorithm based multiple linear regression (GA-MLR). – Fluid Phase Equilibria. – 2013 - №353 – pp15-21
4. I. Paster, M. Shacham, N. Braunerb, Prediction of the Melting Point Temperature Using a Linear QSPR for Homologous Series. – Elsevier. – 2008 – №18 – pp895-900
5. Igor V. Tetko, Daniel M. Lowe, Antony J. Williams, The development of models to predict melting and pyrolysis point data associated with several hundred thousand compounds mined from PATENTS. – J Cheminform. – 2016 – №8(1) – 2:1-2:18
6. Inverse distance weighting [Электронный ресурс]. URL: https://en.wikipedia.org/wiki/Inverse_distance_weighting

7. Modarresi H., Dearden JC, Modarress H, QSPR correlation of melting point for drug compounds based on different sources of molecular descriptors. – J. Chem. Inf. Model. – 2006 - №46 – pp930-936

8. Muthukumarasamy Karthikeyan, Renu Vyas, Practical Chemoinformatics. – Springer. – 2014 – 532p

9. Omar Deeb, Mohammad Goodarzibc, Sherin Alfalah, Prediction of melting point for drug-like compounds via QSPR methods. – Molecular Physics. – 2011 – №4 – pp507-516

10. OPEN Notebook Science [Электронный ресурс]. URL: http://lxsrv7.oru.edu/~alang/meltingpoints/meltingpointof.php?csid=2951

11. Robert C. Glen, Andreas Bender, General Melting Point Prediction Based on a Diverse Compound Data Set and Artificial Neural Networks. – J. Chem. Inf. Model. – 2005 - №45(3) – pp581-590

12. База данных DrugBank [Электронный ресурс]. URL: http://www.drugbank.ca/

13. Введение в R [Электронный ресурс]. URL: http://www.inp.nsk.su/~baldin/DataAnalysis/R/R-01-intro.pdf

14. Выборка Alpha Aesar [Электронный ресурс]. URL: http://usefulchem.blogspot.ru/2011/02/alfa-aesar-melting-point-data-now.html

15. Д. В. Смолин, Введение в искусственный интеллект: конспект лекций. – 2-е изд., перераб. – М.: ФИЗМАТЛИТ – 2007 – 264с.

16. Дескрипторы библиотеки RDKit языка Python [Электронный ресурс]. URL: http://www.RDKit.org/docs/GettingStartedInPython.html#list-of-available-descriptors

17. Конформационные свойства молекул [Электронный ресурс]. URL: http://chemlib.ru/books/item/f00/s00/z0000030/st007.shtml

18. М. Я. Головенко, Физико-химическая фармакология: Монография. – Одесса: Астропринт. – 2004 – 720с.

19. Метод ближайших соседей [Электронный ресурс]. URL: http://www.machinelearning.ru/wiki/index.php?title=Метод_ближайшего_соседа

20. Прогнозирование температуры плавления с помощью CDK-дескрипторов [Электронный ресурс]. URL: http://onschallenge.wikispaces.com/MeltingPointModel001

21. Роберт И. Кабаков, R в действии. Анализ и визуализация данных в языке R / пер. с англ. Полины А. Волковой. – М: ДМК Пресс. – 2014 – 580с.

22. Скользящий контроль [Электронный ресурс]. URL: http://www.machinelearning.ru/wiki/index.php?title=Кросс-валидация

23. Сложная задача прогнозирования температуры плавления [Электронный ресурс]. URL: https://news.brown.edu/articles/2015/07/melting

24. Справочник химика 21 века [Электронный ресурс]. URL: http://chem21.info/page/088211167082149170182133075133154169251076126022/