

КОМПЬЮТЕРНОЕ МАТЕРИАЛОВЕДЕНИЕ ХИМИЧЕСКИХ СОЕДИНЕНИЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ИХ СВОЙСТВ

В.Д. Васильева^{*}, И.В. Гермашев^{}, Е.В. Дербишер^{*},
А.Ю. Александрина^{***}, В.Е. Дербишер^{*}**

^{*}Волгоградский государственный технический университет (ВолгГТУ), Волгоград

^{**}Волгоградский государственный педагогический университет, Волгоград

^{***}Волжский политехнический институт, филиал ВолгГТУ, Волжский

В области материаловедения химических соединений в последние годы продолжается развиваться направление, связанное с прогнозированием физико-химических и технологических свойств соединений по их молекулярной структуре на основе компьютерного моделирования зависимости "структура – свойство – техническая функция" [1,2].

В основу концепции анализа взаимосвязи структуры и свойств легли эмпирические закономерности изменения свойств веществ с изменением их структуры, установленные к середине XIX века. Открытый Д.И. Менделеевым Периодический закон можно рассматривать в качестве первоначальной научной основы концепции "структура – свойство", а предсказания физических свойств и существования неоткрытых новых элементов является примером устойчивой предсказательной модели этой взаимосвязи [3].

При проектировании полимерных композиций, свойства которых в значительной степени определяются разнообразными мономерными добавками, имеющими различные технические функции, большое значение приобретает прогнозирование функциональной активности соединений-добавок в отношении полимерной матрицы на предпроектной стадии создания полимерных композиционных материалов (ПКМ). Основные классы функциональных добавок и проблемы, решаемые с помощью их введения в полимерные композиции, представлены на рисунке, который иллюстрирует почти неисчерпаемое многообразие вариантов.

Данная работа посвящена созданию методики прогнозирования технологической активности низкомолекулярных органических добавок к ПКМ, в частности, к полипропилену (ПП), что обусловлено значительной практической ценностью данного полимера, обладающего при целом комплексе его ценных потребительских свойств и некоторыми недостатками (повышенная склонность к термоокислительной деструкции, низкая светостойкость в атмосферных условиях и др.),

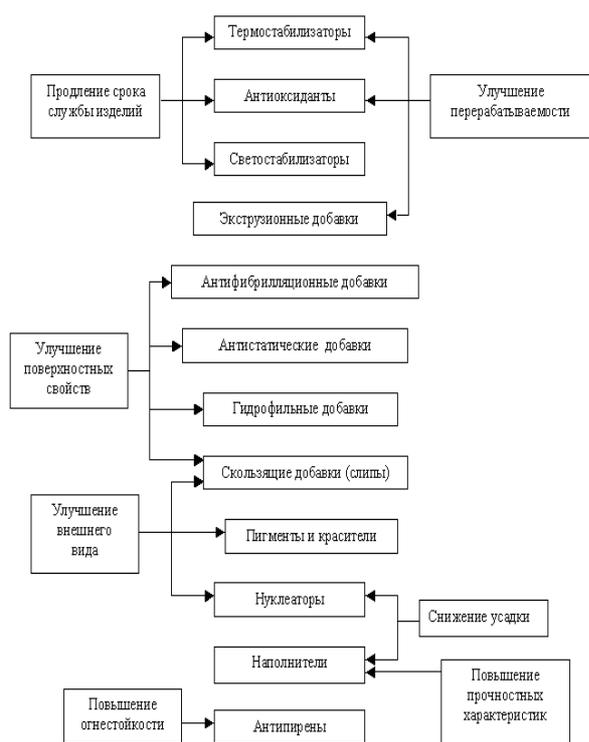
нивелирование которых требует создания полимерных композиций путем введения в полимерную матрицу ряда, в большинстве своем, низкомолекулярных органических добавок. Поэтому представлялось важным создание методической основы для автоматизации прогнозирования свойств активных добавок к ПП на предпроектной стадии получения композиций на его основе, что сократит издержки и период разработки новых добавок и поможет в решении ряда других, например, экологических задач.

Рассматриваемая методика прогнозирования (диагностики) включает следующие последовательные стадии:

1) подготовку и анализ в базе данных (БД) химической информации о соединениях с учетом поставленной задачи и формирования обучающей выборки соединений, достоверно обладающих и не обладающих данным свойством (технической функцией);

Одним из резервов улучшения метрологических характеристик методик АЭС анализа является изменение скорости поступления примесей в зону разряда либо термических условий возбуждения спектров. Чаще всего это достигается введением в пробы носителей, в качестве которых нами исследованы хлориды и карбонаты натрия и свинца. Оценку эффективности действия носителей в присутствии преобладающих количеств кремния проводили с использованием стандартного образца стекла СЛ 96-1. Для этого его разбавляли в соотношении 1:10 графитовым порошком с таким количеством носителя, чтобы его концентрация составляла 3% мас. В расчете на катион соли. На рис. 4 представлено влияние указанных носителей на аналитический сигнал и его относительное стандартное отклонение на примере железа, магния, бария, из которого видно, что соли натрия не выполняют своего оптимизирующего действия.

КОМПЬЮТЕРНОЕ МАТЕРИАЛОВЕДЕНИЕ ХИМИЧЕСКИХ СОЕДИНЕНИЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ИХ СВОЙСТВ



Одним из резервов улучшения метрологических характеристик методик АЭС анализа является изменение скорости поступления примесей в зону разряда либо термических условий возбуждения спектров. Чаще всего это достигается введением в пробы носителей, в качестве которых нами исследованы хлориды и карбонаты натрия и свинца. Оценку эффективности действия носителей в присутствии преобладающих количеств кремния проводили с использованием стандартного образца стекла СЛ 96-1. Для этого его разбавляли в соотношении 1:10 графитовым порошком с таким количеством носителя, чтобы его концентрация составляла 3% мас. В расчете на катион соли. На рис. 4 представлено влияние указанных носителей на аналитический сигнал и его относительное стандартное отклонение на примере железа, магния, бария, из которого видно, что соли натрия не выполняют своего оптимизирующего действия.

Сбор данных для наполнения БД осуществлялся на основе использования справочных материалов, периодических публикаций, из Internet и других источников. В процессе наполнения БД проводилась сортировка соединений-добавок к ПП по функциональному назначению (термостабилизаторы, антиоксиданты, светостабилизаторы,

антипирены и т.д.), а также ранжирование соединений по уровню активности. В результате был организован набор данных (обучающая выборка), содержащий достоверную информацию о 420 органических соединениях; 210 из них образовали группу активных соединений (в т.ч. 130 – высокоактивных и 80 – умеренноактивных), 210 – неактивных. В качестве неактивных в каждом классе ингредиентов определялись соединения, для которых четко отмечена активность только в отношении другой полимерной матрицы (с учетом механизма действия).

Основной информацией, на основе которой формировалось описание химического соединения в БД, являлась его структурная формула. При заполнении БД структура соединения рисовалась с помощью редактора химических структур и фиксировалась в mol-файле, который является составной частью SD-файла (Structure Data File), представляющего базу данных по отдельным структурам в стандартном текстовом формате.

Практическая реализация второй стадии процесса осложнена требованием сохранения максимальной информативности о молекулярной структуре. Одним из современных подходов к количественному описанию структуры химических соединений является дескрипторный подход: описание осуществляется с помощью независимых переменных, характеризующих структурные особенности соединений или их частей – заместителей или фрагментов.

Реализация третьей стадии процесса в современных условиях существенно упростилась в связи с прогрессом вычислительной техники. Построенные статистические модели проходят процедуру оптимизации с последующей генерацией химических структур, имеющих в определенном смысле оптимальное содержание дескрипторов.

Для описания структур соединений-добавок к ПП и вывода решающего правила предложено использовать нейронные сети. Нейронные сети, как правило, оперируют с информацией, заданной в виде чисел, которые подаются на входные синапсы, а ответ также представляется в виде чисел. Поэтому для анализа химических структур необходима, во-первых, особая архитектура нейронной сети и, во-вторых, подходящие под эту архитектуру дескрипторы химической структуры.

Описание структур соединений производилось в виде всевозможных линейных моле-

кулярных цепочек. При этом использовались следующие виды дескрипторов молекулярных цепочек:

1) атомный дескриптор (АД) – атом (кроме водорода) или атом (кроме водорода), объединенный химической связью с одним или несколькими атомами водорода, рассматриваемые как одна целая структурная единица. При этом различают АД для одних и тех же атомов, но соединенных с молекулой разными химическими связями (например, –О–, =О – разные АД). Примерами АД также могут служить –NH₂, –CH=, >P– и т. д.;

2) дескриптор длины (ДД) – число последовательных связей, соединяющих два каких-либо АД. Дескриптор длины принимает значения из множества натуральных чисел;

3) дескриптор связи (ДС) показывает типы связей в молекулярной цепочке: σ-связь, π-связь и химическую связь, отличную от ковалентной.

В качестве примера можно привести разложение на всевозможные линейные молекулярные цепочки структурной формулы молекулы этилового спирта (CH₃-CH₂-OH):

(–CH₃, 1, –CH₂–),
(–CH₃, 1, –CH₂–, 1, –OH),
(–CH₂–, 1, –CH₃),
(–CH₂–, 1, –OH),
(–OH, 1, –CH₂–),
(–OH, 1, –CH₂–, 1, –CH₃).

В цепочке между АД стоит код химической связи, их соединяющей (1 – σ-связь, 2 – двойная π-связь и т. п.).

Принадлежность тестируемого соединения к определенному классу активности определялась с помощью решающего правила. Решающее правило строилось по следующей схеме.

Пусть задан набор веществ $S = \{s_i | i = 1, \dots, n\}$ и m -мерное параметрическое пространство признаков $Q^m = Q_1 \times \dots \times Q_m$, с некоторой метрикой $\rho: Q^m \times Q^m \rightarrow R$, определяющей расстояние между двумя точками пространства Q^m . Каждому веществу s_i поставлен в соответствие вектор значений признаков

$$x^i \in Q^m, i = 1, \dots, n.$$

Также заданы классы эквивалентности $A = \{A_1, \dots, A_k\}$. Для каждого нового вещества необходимо найти класс, к которому он принадлежит. Отнесение к классу производится путем его сравнения с типичными элементами разных классов и выбора ближайшего. В терминах построенного пространства Q^m ма-

лое значение $\rho(x^i, x^j)$ будет означать сходство свойств соответствующих веществ s_i и s_j .

Далее проводится кластерный анализ множества S над параметрическим пространством Q^m [4-6].

Таким образом, накапливаются знания о классах веществ. Для определенности положим: $A_4 = A_3 \cup A_2$ – "активные соединения", A_3 – "высоко активные соединения", A_2 – "умеренно активные соединения", A_1 – "неактивные соединения". Далее на основе накопленной информации проводится обучение распознающей системы (нейронной сети) в следующей последовательности.

1. Из обучающей выборки берется текущее соединение (вещество) и его параметры подаются на входные синапсы обучаемой нейронной сети. В нашем случае – это молекулярные цепочки.

2. Нейронная сеть производит заданное количество тактов функционирования, при этом вектор входных сигналов распространяется по связям между нейронами и модифицирует их, настраивая на увеличение пропускной способности на соответствующих синапсах (если соединение активно) или на уменьшение (если соединение неактивно).

3. Пункты 1 и 2 повторяются для всей обучающей выборки.

4. Формируется тестирующая выборка химических соединений (образцов с известными прогнозируемыми параметрами (техническими функциями).

5. Подаются на входные синапсы параметры (молекулярные цепочки) химического соединения из тестирующей выборки.

6. Измеряются сигналы выходных нейронов.

7. Производится интерпретация выданных сигналов и вычисляется оценка, характеризующая различие между выданным сетью ответом и требуемым ответом, имеющимся в образце. Чем меньше оценка, тем лучше распознан образец.

8. Если оценка образца недостаточно мала, вычисляются поправочные коэффициенты для каждого синаптического веса матрицы связей, после чего проводится подстройка синаптических весов.

9. Осуществляется переход к следующему образцу (п. 5).

Таким образом, предлагаемая методика позволяет не только оценивать свойства веществ внутри БД, но и осуществлять прогнозирование в других наборах; тестировать новые химические структуры на проявление

КОМПЬЮТЕРНОЕ МАТЕРИАЛОВЕДЕНИЕ ХИМИЧЕСКИХ СОЕДИНЕНИЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ИХ СВОЙСТВ

функциональной активности в отношении выбранной полимерной матрицы (в данном случае ПП) и отбирать соединения для экспериментального подтверждения прогнозируемых свойств, тем самым резко ограничив пространство целенаправленного поиска эффективных добавок к полимерам.

СПИСОК ЛИТЕРАТУРЫ:

1. Дербишер В.Е., Гермашев И.В., Васильев П.М. Компьютеризированная методика прогнозирования активных добавок к полимерным композициям // Пластические массы. – 1999. – № 2. – С. 32-36.
2. Орлов В.В., Дербишер В.Е., Зотов Ю.Л., Васильев П.М., Гермашев И.В., Дербишер Е.В., Колоскова А.Ю. Диагностика возможной активности производных адамантана в полимерных композициях методами молекулярного дизайна // Химическая промышленность. – 2003. – Т. 80, № 2. – С. 46-55.
3. Раевский О.А. Дескрипторы молекулярной структуры в компьютерном дизайне биологически активных веществ // Успехи химии. – 1999. – Т. 68. Вып. 6. – С. 555-575.
4. Факторный, дискриминантный и кластерный анализ. – М.: Финансы и статистика, 1989. – 215 с.
5. Гермашев И.В., Дербишер В.Е., Васильев П.М. Прогнозирование на основе вероятностных методов активности низкомолекулярных органических соединений в полимерных композициях // Теоретические основы химической технологии. – 1998. – Т. 32, № 5. – С. 563-567.
6. Тюрин Ю.Н., Макарова А.А. Анализ данных на компьютере. – М.: ИНФРА-М, 2003. – 544 с.