

УДК 539.189.1

ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКОГО АЛГОРИТМА ДЛЯ ОБРАБОТКИ ИК-ФУРЬЕ-СПЕКТРОВ ПОЛИМЕРОВ

*Э.Р. Шаймухаметова, Д.З. Галимуллин,
М.Э. Сибгатуллин, Д.И. Камалова, М.Х. Салахов*

Аннотация

Показана возможность использования генетического алгоритма для решения проблемы разложения сложных молекулярных спектров на составляющие. На примере модельных спектров исследовано качество восстановления спектра в зависимости от уровня шума, полуширины, интенсивности и положения компонент.

Ключевые слова: генетический алгоритм, обработка спектров, ИК-фурье-спектры полимеров.

Введение

Колебательная спектроскопия позволяет получать новые данные о топологической структуре и молекулярной подвижности разветвленных полимеров, устойчивый интерес к которым наблюдается в последние десятилетия [1]. Локальная молекулярная подвижность определяет многие физические и химические свойства полимеров, поэтому к настоящему времени предложен целый ряд физических (экспериментальных и теоретических) методов ее исследования и разработка новых методов продолжается. Однако инфракрасная спектроскопия до сих пор продолжает оставаться наиболее распространенным физическим методом исследования полимеров. Такое значение ИК-спектроскопии объясняется тем, что качественная интерпретация спектра достаточно проста и не требует от исследователя специальной физической и математической подготовки. Но в практике спектроскопических исследований часто имеют дело со сложными спектральными контурами, имеющими неразрешенную структуру, и спектры поглощения многоатомных органических соединений, в частности полимеров, характерны в этом отношении. Необходимую исследователю информацию можно получить, только определив параметры отдельных полос, входящих в состав сложного спектра. В связи с этим разложение сложного контура на составляющие продолжает оставаться важной задачей прикладной спектроскопии [2]. В настоящей статье приводятся результаты применения нового подхода к решению проблемы разложения сложных спектральных контуров на составляющие, а именно генетического алгоритма, приспособленного для интерпретации инфракрасных полос поглощения макромолекул разветвленных полиметилметакрилатов с целью получения информации о локальной молекулярной динамике.

1. Общие сведения о генетическом алгоритме

Преимуществом генетических алгоритмов является возможность их применения не только для решения сложных задач, для которых не существует других

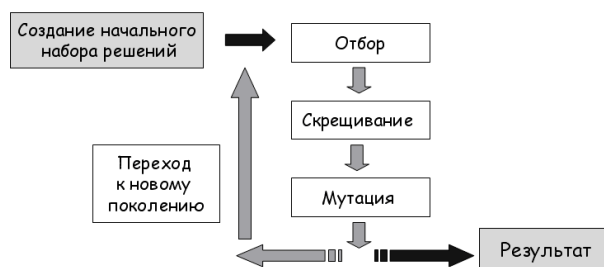


Рис. 1. Общая схема генетического алгоритма

специальных методов, но и в случаях, когда существующие методы работают, но есть возможность достигнуть улучшения путем комбинирования их с генетическими алгоритмами. Генетический алгоритм – это метод поиска и оптимизации решения многопараметрической задачи, работающий аналогично процессу эволюции в природе, то есть использующий так называемые «операторы эволюции»: отбор (селекцию), мутацию, скрещивание (кроссовер). Генетический алгоритм решает проблему, генерируя массу случайных программ – вариантов решения проблемы, запуская и оценивая каждый из них согласно определенным критериям приспособленности. Генетический алгоритм выбирает из каждого «поколения» лучшие варианты решения и получает от них «потомство», что аналогично естественному отбору в природе. На рис. 1 представлена общая схема генетического алгоритма [3].

При реализации генетического алгоритма [3] в качестве первого шага следует случайным образом создать некую начальную «популяцию», при этом если даже она окажется совершенно неконкурентоспособной, генетический алгоритм все равно достаточно быстро переведет ее в жизнеспособную «популяцию». При выполнении «селекции» применяют оператор, посредством которого выбираются решения для «скрещивания» и получения «потомков». Для имитации естественной селекции решения с более высокой пригодностью должны выбираться с большей вероятностью. «Скрещивание» использует оператор, за счет которого формируется следующий набор решений – «потомство». Главное требование к оператору «скрещивания» состоит в том, чтобы «потомок» или «потомки» имели возможность «унаследовать» черты обоих (или всех) «родителей», смешав их каким-либо способом. «Мутация» заключается в выполнении небольших изменений в значениях одного или нескольких параметров решения. В генетических алгоритмах «мутация» рассматривается как восстановление потерянного генетического материала, но не поиск лучшего решения.

Выделяют два типа формирования нового «поколения» после получения множества «детей» в результате «кроссовера» и «мутации»: 1) «дети» замещают «родителей»; 2) новое «поколение» составляется из совокупности и «детей», и их «родителей», например, выбором N лучших [4]. Для формирования нового «поколения» возможно также использование принципа элитизма, когда в новое «поколение» обязательно включается заданное количество лучших «особей» предыдущего «поколения» (часто одна лучшая «особь»). Использование второй стратегии и элитизма оказывается весьма полезным для повышения эффективности генетического алгоритма, так как не допускает потерю лучших решений.

Согласно теореме шаблонов при «популяциях» со средним размером [5] генетический алгоритм следует разрабатывать с учетом в равной мере влияния операторов «мутации» и «кроссовера». Принимая это во внимание, мы разработали новый вариант использования генетического алгоритма для математической обработки

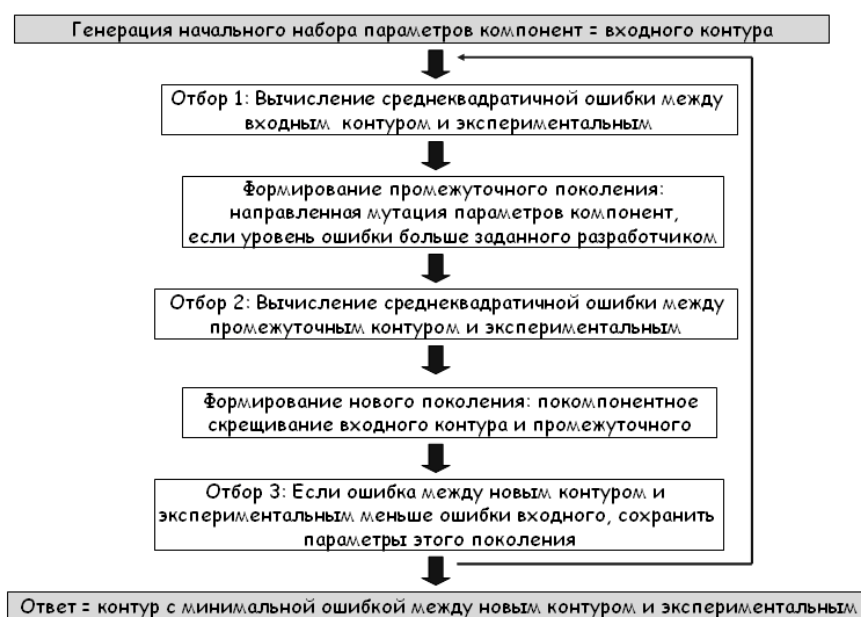


Рис. 2. Генетический алгоритм для обработки спектральных данных

ИК-фурье-спектров, при этом последовательность действий состоит в следующем (рис. 2).

Создается начальный набор параметров составляющих спектра, вводятся их количество и положение центров. На первом этапе производится оценка ошибки между входным контуром и экспериментальным в каждом интервале, центром которого является элементарная составляющая спектра. Далее с использованием оператора направленной «мутации» производится формирование промежуточного «поколения».

На этапе второго отбора снова вычисляется покомпонентная ошибка на каждом интервале, но уже между промежуточным и входным контурами. Применением оператора «скрещивания» к промежуточному и входному контурам получают новый контур с наименьшей разницей между восстановленным и реальным спектрами. Сохраняя параметры лучшего набора в процессе каждой итерации, в конце цикла выбирают наилучший набор (с наименьшей ошибкой восстановления) и выводят его как ответ. При наличии нескольких локальных экстремумов такой тип генетических алгоритмов позволяет выбрать наилучший вариант решения.

2. Отработка методики на модельных контурах

Сначала мы исследовали способность генетического алгоритма разделять сложные зашумленные модельные спектры. Для исследования были использованы дублет, триплет и контур, построенный из шести составляющих. На модельный сигнал накладывался шум с нормальным распределением в интервале от 0 до 20%. Исследования показали, что при вычислении ошибки между восстановленным контуром и модельным контуром без шума для всех исследуемых модельных спектров ошибка плавно растет пропорционально шуму, который для ИК-спектров обычно составляет около 3%. Как показано на рис. 3, ошибка восстановления контура генетическим алгоритмом при таком уровне шума составляет около 1.5%, то есть генетический алгоритм может быть использован для разделения конту-

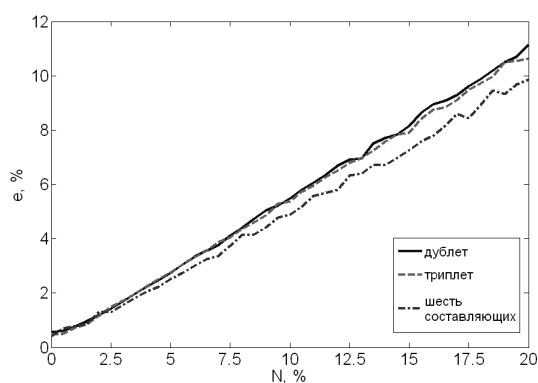


Рис. 3. Относительная ошибка восстановления (e) сложного контура в зависимости от уровня шума (N)

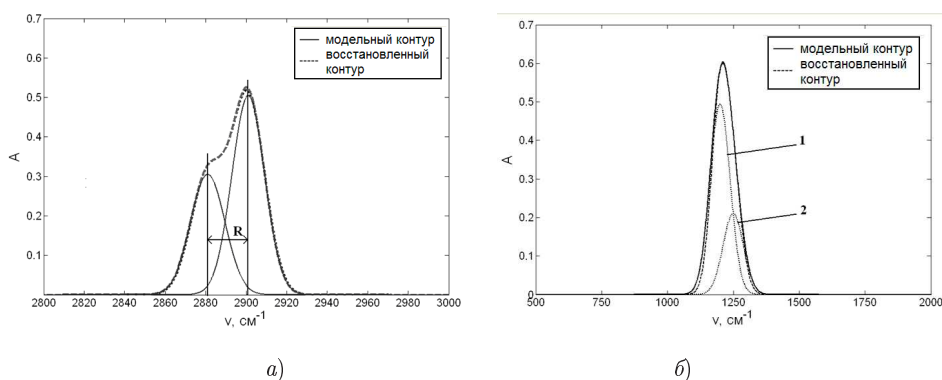


Рис. 4. (а) модельный дублет, состоящий из одинаковых по полуширине, но различных по интенсивности полос; (б) модельный дублет, состоящий из одинаковых по полуширине, но различных по интенсивности полос. Положение полос, их полуширина и интенсивность полосы 1 оставались неизменными, а интенсивность полосы 2 возрастала

ров в случае слабозашумленных сигналов, например, таких, какими являются сигналы в молекулярной спектроскопии. Исследуемую зависимость в указанном интервале можно аппроксимировать линейной, и тогда коэффициент пропорциональности между уровнем накладываемого шума и уровнем ошибки восстановления составляет для дублета 0.53 ± 0.005 , для триплета – 0.52 ± 0.005 и для контура, состоящего из шести полос, – 0.50 ± 0.005 .

Далее была рассмотрена зависимость точности восстановления контура от расстояния между его составляющими. Мы использовали в качестве модельного контура дублет, состоящий из одинаковых по полуширине, но различных по интенсивности составляющих (рис. 4, а). Положение одной из составляющих оставалось неизменным, а вторую перемещали относительно первой. На рис. 5 приведена зависимость суммарной ошибки восстановления дублета от расстояния между составляющими, выраженного в полуширине полос дублета. В этом случае максимальная ошибка (0.5%) наблюдается при расстоянии между составляющими дублета, равном 1.5 полуширины. Начиная с расстояния, равного 2.5 полуширины полос дублета, исследуемая зависимость выходит на плато при суммарной ошибке, не превышающей 0.4%.

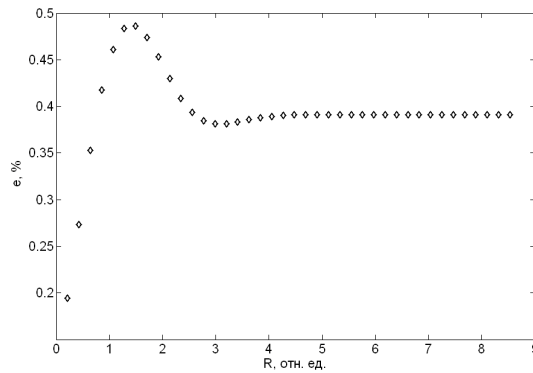


Рис. 5. Зависимость суммарной ошибки (e) восстановления дублета от расстояния (R) между составляющими, выраженного в величинах полуширины полос дублета

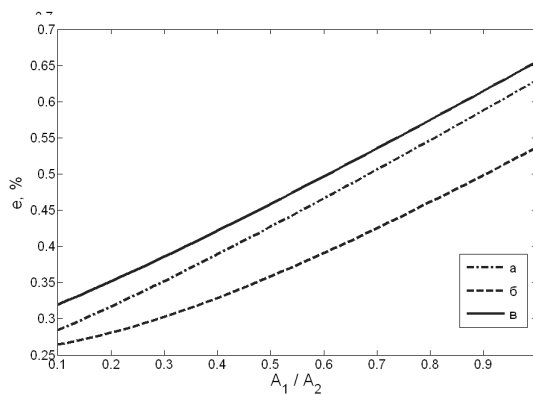


Рис. 6. Зависимость среднеквадратичной ошибки (e) восстановления суммарного контура от отношения интенсивностей полос A_1/A_2 : а) полуширина 12 см^{-1} , расстояние между полосами 1.75 см^{-1} ; б) полуширина 12 см^{-1} , расстояние между полосами 2.5 см^{-1} ; в) полуширина 14 см^{-1} , расстояние между полосами 2.25 см^{-1}

На следующем этапе рассматривалась эффективность восстановления с помощью генетического алгоритма контура, состоящего из сильно отличающихся по интенсивности составляющих полос. Для определенности был взят дублет, состоящий из одинаковых по полуширине, но различных по интенсивности составляющих при постоянном расстоянии между ними (рис. 4, б). Интенсивность одной из полос оставалась постоянной, а у второй изменялась; в результате была рассчитана среднеквадратичная ошибка восстановления суммарного контура. Зависимость ошибки восстановления контура от соотношения интенсивностей составляющих приведена на рис. 6, а. Далее была рассчитана ошибка восстановления суммарного контура при такой же (12 усл. ед.) полуширине, но большем расстоянии между составляющими. В первом случае расстояние составляло 15% от полуширины полос, а во втором – 20%. Зависимость ошибки восстановления при таких условиях приведена на рис. 6, б. После этого была рассчитана ошибка восстановления контура для дублета, состоящего из полос с большей полушириной (14 усл. ед), но расположенных на том же расстоянии относительно друг друга (15% от полуширины полосы) (рис. 6, в).

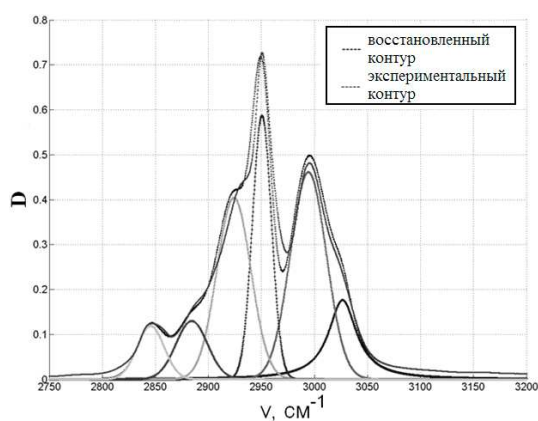


Рис. 7. Восстановление спектра разветвленного сополимера метилметакрилата с помощью генетического алгоритма

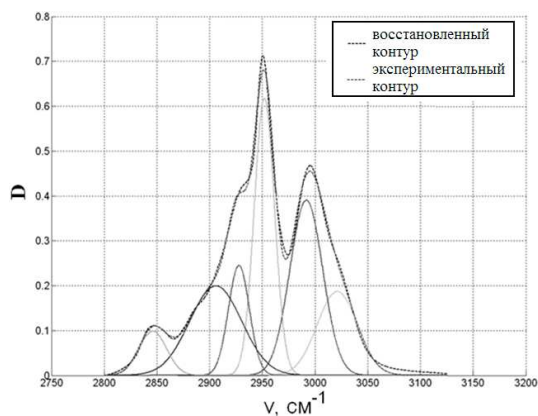


Рис. 8. Восстановление спектра разветвленного сополимера метилметакрилата с помощью метода наименьших квадратов

3. Сравнение генетического алгоритма с методом наименьших квадратов на примере ИК-фурье-спектров полимеров

Исследуемые ИК-фурье-спектры разветвленных сополимеров метилметакрилата в диапазоне $2750 \div 3200 \text{ см}^{-1}$ аппроксимировали с помощью генетического алгоритма. Поскольку экспериментальный контур является сложным (рис. 7) и наблюдается значительное перекрывание его составляющих, необходимо ввести априорную информацию о типе спектральной полосы, а именно о количестве составляющих, их положении и форме.

Для определения количества составляющих спектра мы использовали вейвлет Гаусса второго порядка. Методами вейвлет-производной спектрометрии [6] было установлено, что каждый из сложных спектров поглощения исследуемых сополимеров метилметакрилата в области $2750 \div 3200 \text{ см}^{-1}$ состоит из шести полос, положение центров которых вычислялось по минимумам вейвлет-производной. На рис. 8 показано восстановление спектра сополимера метилметакрилата по шести составляющим, которые можно отнести к симметричным и антисимметричным колебаниям CH_3 - и CH_2 -групп [7]. Отметим, что разложение на составляющие было проведено для ИК-спектров сополимеров в температурном интервале.

Исследуемые ИК-фурье-спектры разветвленных сополимеров метилметакрилата мы для сравнения также восстановили методом наименьших квадратов (рис. 8), который, на первый взгляд, восстанавливает контур с меньшей невязкой, чем генетический алгоритм. Однако метод наименьших квадратов не «чувствует» перегиб в области 2885 см^{-1} , хотя исследование температурной динамики составляющих полос показало, что полоса 2885 см^{-1} при температуре 170 К становится разрешенной. Следовательно, метод наименьших квадратов подгоняет контур, а не разделяет его, минимизируя невязку. В пользу восстановления экспериментального контура с помощью генетического алгоритма указывает и то, что составляющие в этой области, по-видимому, должны иметь не сильно отличающиеся значения полуширины, так как эти полосы поглощения близки по механизму уширения.

Работа выполнена при поддержке гранта Президента РФ НШ-5289.2010.2 и Федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009–2013 годы (ГК № 02.740.11.0428).

Summary

E.R. Shaimukhametova, D.Z. Galimullin, M.E. Sibgatullin, D.I. Kamalova, M.Kh. Salakhov. The Use of Genetic Algorithm for Processing FTIR Spectra of Polymers.

A possibility of resolution enhancement of composite molecular spectra using the genetic algorithm is shown. The quality of spectrum reconstruction depending on the noise level, halfwidth, intensity, and the position of components has been investigated for model spectra.

Key words: genetic algorithm, spectra processing, FTIR spectra of polymers.

Литература

1. Курмаз С.В., Грачев В.П., Кочнева И.С., Перепелицина Е.О., Эстрина Г.А. Синтез, структура, свойства разветвленных полиметакрилатов // Высокомолекул. соед. – 2007. – Т. 49, № 8. – С. 1480–1493.
2. Севастьянов А.А., Харинцев С.С., Салахов М.Х. Разделение сложных спектров с помощью вейвлет-производной спектроскопии // Структура и динамика молекулярных систем: Сб. ст. – Казань: Казан. гос. ун-т, 2003. – С. 277–282.
3. Вороновский Г.К., Махотило К.В., Петрашев С.Н., Сергеев С.А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. – Харьков: ОСНОВА, 1997. – 112 с.
4. Kim D.H., Abraham A., Cho J.H. A hybrid genetic algorithm and bacterial foraging approach for global optimization // Information Sciences. – 2007. – V. 177, No 18. – P. 3918–3937.
5. Lakhmi C.J., Martin N.M. Fusion of Neural Networks, Fuzzy Sets, and Genetic Algorithms: Industrial Applications. – Boca Raton, FL, USA: CRC Press, Inc., 1999. – 354 p.
6. Kharintsev S.S., Kamalova D.I., Salakhov M.Kh., Sevastianov A.A. Resolution enhancement of composite spectra using wavelet-based derivative spectrometry // Spectrochim. Acta A. Mol. Biomol. Spectrosc. – 2005. – V. 61, No 1–2. – P. 149–156.
7. Дехант И. Инфракрасная спектроскопия полимеров / Под ред. И. Деханта. – М.: Химия, 1976. – 472 с.

Поступила в редакцию
05.01.10

Шаймухаметова Эльвира Рамилевна – аспирант кафедры оптики и нанофотоники Казанского (Приволжского) федерального университета.

E-mail: *E-mail: eshaimukhametova@mail.ru*

Галимуллин Дамир Зиннурович – кандидат физико-математических наук, ассистент кафедры оптики и нанофотоники Казанского (Приволжского) федерального университета.

Сибгатуллин Мансур Эмерович – кандидат физико-математических наук, ассистент кафедры оптики и нанофотоники Казанского (Приволжского) федерального университета.

Камалова Дина Илевна – доктор физико-математических наук, профессор кафедры оптики и нанофотоники Казанского (Приволжского) федерального университета.

Салахов Мяззюм Халимулович – доктор физико-математических наук, профессор, заведующий кафедрой оптики и нанофотоники Казанского (Приволжского) федерального университета.