АЛГОРИТМЫ АНАЛИЗА СТРУКТУРНОЙ ИНФОРМАЦИИ
(Вычислительные системы)
1984 год Выпуск 103

УДК 621.43

ОПРЕДЕЛЕНИЕ ИНДЕКСОВ СВЯЗАННОСТИ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ ЦИКЛИЧЕСКОЙ СТРУКТУРЫ С ИСПОЛЬЗОВАНИЕМ ТЕОРИИ ГРАФОВ

И.В.Дылевский, Л.В.Дылевская, З.Ю.Булычева, В.Ф.Кутенёв

Индексы связанности [1,2] являются количественными характеристиками структурной формулы молекулы. Для различных классов химических соединений (например, полициклических ароматических углеводородов) возможно установление функциональных связей между индексами связанности и некоторыми физическими свойствами соединений (такими как температура кипения, молекулярная масса, параметры удерживания в газо-жидкой хроматографии и другими).

Известно, что корреляция экспериментальных структурных параметров (индексов удержания) с индексами связанности первого порядка для сложных циклических соединений имеет достаточно большую величину стандартных отклонений [3]. Для уменьшения этих отклонений был произведен расчет индексов связанности высоких порядков, более полно описывающих исследуемые структуры.

Введение индексов связанности высоких порядков в многофакторные уравнения связи свойства – структура повышает значимость этих уравнений.

Целью настоящей работы является вывод основных теоретических закономерностей, являющихся основой для составления программы вычисления индексов связанности высоких порядков для органических соединений циклической структуры.

Структурная формула молекулы может быть описана с помощью неориентированного графа со взвешенными ребрами $G = (V, E)$, где $V=\{v_1,\dots,v_n\}$ – множество атомов, $E=\{e_1,\dots,e_m\}$ – множество валентных связей. Ребро e_i есть неупорядоченная пара вершин $\{v_{\alpha(i)}, v_{\beta(i)}\}$. Вес c_i ребра e_i – натуральное число, равное валентности химической связи по данному ребру. Из построения следует, что граф G является связным.

Для всех вершин v_j, $j = 1,2,\dots,n$, определяются валентности атомов, находящихся в этих вершинах:

$$Z_j = \sum_{i | v_j \in e_i} c_i \, .$$

В терминах введенных обозначений индекс связанности порядка k определяется по формуле:

$$\chi_k = \sum_{t=\{i_1,\dots,i_{k+1}\}} \left(\prod_{j=1}^{k+1} Z_{i_j} \right)^{-\frac{1}{2}} , \qquad (1)$$

где сумма берется по всем связным подграфам t графа G, содержащим ровно $(k+1)$ вершину. Связный подграф понимается как непустое подмножество t множества V такое, что либо t содержит одну вершину графа G, либо t содержит более одной вершины графа G, и между любыми двумя вершинами, лежащими в t, существует маршрут в графе G, не выводящий за пределы множества t.

Так как необходимо вычислять индексы порядков $0,1,\dots,n-1$, то возникает задача нахождения множества T всех связных подграфов графа G. Для решения этой задачи рассмотрим реберный граф $\tilde{G}$, соответствующий исходному графу, $\tilde{G} = (\tilde{V},\tilde{E})$, где $\tilde{V} = \{e_1,\dots,e_m\}$, т.е. множество вершин реберного графа совпадает с множеством ребер исходного графа, $\tilde{E} = \{\tilde{e}_i\}$ - множество ребер реберного графа, которое строится по следующему правилу: вершины e_i и e_j графа $\tilde{G}$ смежны только в том случае, если ребра e_i и e_j графа G имеют общую вершину. Реберный граф $\tilde{G}$ также является неориентированным.

ЛЕММА I. Пусть $\tilde{t} = \{e_{j_1},\dots,e_{j_p}\}$ - связный подграф в G, $t = \bigcup_{i=1}^{p} \{v_{\alpha(j_i)}, v_{\beta(j_i)}\}$ - индуцированный подграф в G. Тогда t - связный подграф в G.

ДОКАЗАТЕЛЬСТВО. Пусть v_{i_1} и v_{i_2} - любые две вершины, лежащие в t. Если существует ребро в E, соединяющее эти вершины, то оно и является требуемым маршрутом. Если же такого ребра не существует, то, по определению множества t, существуют по крайней мере 2 ребра e_{s_1}, e_{s_2}, лежащие в $\tilde{t}$, имеющие концевыми вершинами v_{i_1} и v_{i_2} соответственно. Поскольку $\tilde{t}$ - связный подграф в $\tilde{G}$, то существует маршрут, соединяющий e_{s_1} и e_{s_2} в $\tilde{G}$ и не выводящий за пределы множества $\tilde{t}$. Но тогда последовательность вершин из $\tilde{t}$, через которые проходит этот маршрут, является маршрутом в G, соединяющим вершины v_{i_1} и v_{i_2} и не выводящим за пределы множества t. Лемма доказана.

ЛЕММА 2. Пусть $t = \{v_{i_1}, \dots, v_{i_s}\}$ - связный подграф в G, состоящий не менее чем из двух вершин. Тогда ему соответствует некоторый связный подграф $\tilde{t} = \{e_{j_1}, \dots, e_{j_r}\}$ в $\tilde{G}$ (не обязательно единственный) такой, что $\bigcup_{i=1}^{r} \{v_{\alpha(j_i)}, v_{\beta(j_i)}\} = t$.

ДОКАЗАТЕЛЬСТВО. Так как t - связный подграф в G, то существуют маршруты, соединяющие всевозможные пары вершин из t, не выводящие за пределы множества t. Рассмотрим подграф $\tilde{t} = \{e_{j_1}, \dots, e_{j_r}\}$, состоящий из объединения всех таких маршрутов. Тогда $\bigcup_{i=1}^{r} \{v_{\alpha(j_i)}, v_{\beta(j_i)}\} \subseteq t$. Но так как каждая вершина из t находится хотя бы в одном из рассмотренных маршрутов, то $t \subseteq \bigcup_{i=1}^{r} \{v_{\alpha(j_i)}, v_{\beta(j_i)}\}$, т.е. $\bigcup_{i=1}^{r} \{v_{\alpha(j_i)}, v_{\beta(j_i)}\} = t$.

Докажем связность подграфа $\tilde{t}$. Возьмем любые две вершины e_p, e_q подграфа $\tilde{t}$. Рассмотрим их концевые вершины $v_{\alpha(p)}$, $v_{\alpha(q)}$. Из вышесказанного $v_{\alpha(p)} \in t$, $v_{\alpha(q)} \in t$. Если $v_{\alpha(p)} = v_{\alpha(q)}$, то по определению $\tilde{E}$ вершины e_p и e_q являются смежными в $\tilde{G}$, и наличие маршрута в $\tilde{G}$, соединяющего e_p и e_q, доказано.

Если же $v_{\alpha(p)} \neq v_{\alpha(q)}$, то в G существует маршрут, соединяющий вершины $v_{\alpha(p)}$ и $v_{\alpha(q)}$, не выводящий за пределы t. Если к концевым ребрам этого маршрута добавить ребра e_p и e_q, то снова получится маршрут в G, не выводящий за пределы t. Так как любые два соседних ребра маршрута, рассматриваемые как вершины $\tilde{G}$, являются смежными, то получается маршрут в $\tilde{G}$, соединяющий вершины e_p и e_q и не выводящий за пределы множества $\tilde{t}$. Лемма доказана.

Для определения множества T всех связных подграфов графа воспользуемся реберным графом $\tilde{G}$.

Пусть $\tilde{T}$ - множество всех связных подграфов реберного графа $\tilde{G}$. Определим множество подграфов исходного графа, индуцируемых связными подграфами реберного графа:

$$A = \{\bigcup_{i=1}^{p} \{v_{\alpha(j_i)}, v_{\beta(j_i)}\} | \{e_{j_1}, \dots, e_{j_p}\} \in \tilde{T}\}. \tag{2}$$

Обозначим

$$B = \{\{v_1\}, \{v_2\}, \dots \{v_n\}\}. \tag{3}$$

Тогда верна следующая

ТЕОРЕМА I. $T = A \cup B$.

ДОКАЗАТЕЛЬСТВО. Пусть $t \in T$. Если t содержит ровно одну вершину, то $t \in B$, если же содержит не менее двух вершин, то, по лемме 2, $t \in A$. Следовательно, $t \in A \cup B$ и $T \subseteq A \cup B$.

Обратно, если $t \in B$, то $t \in T$. Если же $t \in A$, то соответствующий подграф $\tilde{t} = \{e_{j_1}, \dots, e_{j_p}\}$, связный в $\tilde{G}$ и, по лемме I, подграф t будет связным в G, т.е. $t \in T$. Следовательно, $A \cup B \subseteq T$ и $T = A \cup B$. Теорема доказана.

Результатом доказанной теоремы является следующее: для определения всех индексов связанности по формулам (I) достаточно найти множество всех связных подграфов реберного графа и по соотношениям (2), (3) перейти к множеству всех связных подграфов исходного графа.

Опишем индуктивный метод получения всех связных подграфов реберного графа. Введем следующие обозначения:

$$\tilde{V}^{(1)} = \tilde{V}, \quad \tilde{E}^{(1)} = \tilde{E}; \quad \tilde{G}^{(1)} = (\tilde{V}^{(1)}, \tilde{E}^{(1)});$$

$$\tilde{V}^{(k)} = \tilde{V}^{(k-1)} \setminus \{e_{k-1}\}, \quad \tilde{E}^{(k)} = \tilde{E}^{(k-1)} \setminus \{\tilde{e}_i | e_{k-1} \in \tilde{e}_i\},$$

$$\tilde{G}^{(k)} = (\tilde{V}^{(k)}, \tilde{E}^{(k)}), \quad k = 2, \dots m .$$

Граф $\tilde{G}^{(2)}$ получается из графа $\tilde{G}^{(1)}$ отбрасыванием вершины e_1 и всех ребер $\tilde{e}_i$, инцидентных вершине e_1, и т.д.

Обозначим через $\tilde{T}^{(k)}$ множество всех связных подграфов графа $\tilde{G}^{(k)}$. Так как $\tilde{G}^{(1)} = \tilde{G}$, то $\tilde{T}^{(1)} = \tilde{T}$. Граф $\tilde{G}^{(m)}$ имеет простую структуру: $\tilde{V}^{(m)} = \{e_m\}$, $\tilde{E}^{(m)} = \emptyset$. Для него $\tilde{T}^{(m)} = \{\{e_m\}\}$. Для получения рекуррентного соотношения между множествами $\tilde{T}^{(k+1)}$ и $\tilde{T}^{(k)}$ докажем следующую лемму.

ЛЕММА 3. *Любая вершина связного подграфа $\tilde{t}$ графа $\tilde{G}^{(k)}$ разбивает его не более чем на два отдельных связных подграфа.*

ДОКАЗАТЕЛЬСТВО. Пусть $\mathfrak{T} = \{e_{i_1}, \dots e_{i_s}, \dots e_{i_l}\}$. Рассмотрим подграф $\mathfrak{T}$ без вершины e_{i_s}. Пусть он распадается более чем на два связных подграфа $\mathfrak{T}_1 = \{e_{p_1}, \dots, e_{p_u}\}$, $\mathfrak{T}_2 = \{e_{q_1}, \dots, e_{q_v}\}$, $\mathfrak{T}_3 = \{e_{r_1}, \dots, e_{r_w}\}$ и т.д. Так как подграфы $\mathfrak{T}_1, \mathfrak{T}_2, \mathfrak{T}_3$ не связаны один с другим, то в множестве $\tilde{E}^{(k)}$ нет ребер, соединяющих какие-либо две вершины разных подграфов. Следовательно, в единый связный подграф подграфы $\mathfrak{T}_1, \mathfrak{T}_2, \mathfrak{T}_3$ объединяет вершина e_{i_s}, т.е. существуют вершины $e_{j_1} \in \mathfrak{T}_1$, $e_{j_2} \in \mathfrak{T}_2$, $e_{j_3} \in \mathfrak{T}_3$ такие, что e_{i_s} смежна с вершинами $e_{j_1}, e_{j_2}, e_{j_3}$ в $\tilde{G}^{(k)}$. Но так как вершина e_{i_s} есть ребро $e_{i_s} = \{v_{\alpha(i_s)}, v_{\beta(i_s)}\}$ исходного графа и она может быть смежной с вершиной e_{j_1}, только если у ребер e_{i_s} и e_{j_1} есть общая вершина v_{γ_1} в исходном графе. Аналогично у ребер e_{i_s} и e_{j_2} есть общая вершина v_{γ_2}, у ребер e_{i_s} и e_{j_3} – общая вершина v_{γ_3}. Так как ребро имеет только две вершины, то по крайней мере две вершины из трех $v_{\gamma_1}, v_{\gamma_2}, v_{\gamma_3}$ совпадают. Без ограничения общности можно считать, что $v_{\gamma_1} = v_{\gamma_2}$. Но тогда e_{j_1} и e_{j_2} имеют общую вершину $v_{\gamma_1} = v_{\gamma_2}$, поэтому в $\tilde{E}^{(k)}$ должно существовать ребро, соединяющее вершины e_{j_1} и e_{j_2}. Полученное противоречие доказывает лемму.

ТЕОРЕМА 2. *Множество $\tilde{T}^{(k)}$ состоит из объединения четырех множеств:*

I) множества $\tilde{T}^{(k+1)}$;

2) множества $\{e_k\}$;

3) всех объединений вершины e_k и тех множеств из $\tilde{T}^{(k+1)}$, в которых присутствует хотя бы одна вершина, смежная с вершиной e_k;

4) всех объединений вершины e_k и двух множеств из $\tilde{T}^{(k+1)}$ таких, что в каждом из них есть хотя бы по одной вершине, смежной с вершиной e_k.

ДОКАЗАТЕЛЬСТВО. Множество $\tilde{T}^{(k)}$ можно представить как объединение непересекающихся множеств $\tilde{T}^{(k)} = \tilde{T}^{(k+1)} \cup A$, где $\tilde{T}^{(k+1)}$ – множество связных подграфов в $\tilde{G}^{(k+1)}$, следовательно, $\tilde{T}^{(k+1)}$ есть множество связных подграфов в $\tilde{G}^{(k)}$, не содержащих вершину e_k; A - множество связных подграфов в $\tilde{G}^{(k)}$, содержащих вершину e_k.

Но так как по лемме 3 вершина e_k разбивает любой связный подграф в $\tilde{G}^{(k)}$ не более чем на два связных подграфа, не содержащих вершину e_k (т.е. подграфы из $\tilde{T}^{(k+1)}$), то элементы из A получаются либо путем присоединения ко всем подграфам из $\tilde{T}^{(k+1)}$ вершины e_k при условии сохранения связности полученного подграфа в $\tilde{G}^{(k)}$, либо путем объединения двух связных подграфов из $\tilde{T}^{(k+1)}$ с помощью вершины e_k в один связный подграф в $\tilde{G}^{(k)}$. Кроме этого, A не содержит элемента, состоящего из одной вершины e_k. Следовательно, A есть объединение множеств, указанных в пп. 2-4. Теорема доказана.

Переход от исходного графа к рассмотрению реберного графа при вычислении индексов связанности обусловлен требованием упрощения машинной реализации алгоритма вычисления индексов связанности. Простая процедура получения множеств $\tilde{T}^{(k)}$, $k = m-1, \ldots, 1$, связных подграфов получается только благодаря выполнению леммы 3 для реберного графа. Для исходного графа лемма 3 неверна, что легко проверяется на примере графа $G = (V, E)$, где $V = \{v_1, v_2, v_3, v_4\}$, $E = \{\{v_1, v_4\}, \{v_2, v_4\}, \{v_3, v_4\}\}$. Вершина v_4 разбивает связный подграф $\{v_1, v_2, v_3, v_4\}$ на три связных подграфа $\{v_1\}, \{v_2\}, \{v_3\}$.

Литература

I. RANDIČ M. On Characterization of Molecular Branching. - J. Americ.Chem.Soc.,1975,v.97,N 23,p.6609-6615.

2. KIER L.B., HALL L.H. Molecular Connectivity in Chemistry and Drug Research. - N.Y.: Academic Press,1976,v.14.

3. KALISZAN R., LAMPARCZYC H. Relationship between Molecular Connectivity and Retention Indices for Polycyclic Aromatic Hydrocarbons.- J.Chromatogr.Sci.,1978,v.16,p.246-248.

Поступила в ред.-изд.отд.
27 декабря 1983 года