

КРАТКИЕ СООБЩЕНИЯ

УДК 543.42 : 002.6+681.32

М. И. ГОРФИНКЕЛЬ, С. А. НЕХОРОШЕВ
 (Новосибирск)

**ОБ ИСПОЛЬЗОВАНИИ НЕКОТОРЫХ ОБОБЩЕННЫХ ХАРАКТЕРИСТИК
 ДЛЯ СЖАТИЯ СПЕКТРАЛЬНОЙ ИНФОРМАЦИИ**

За последние годы получили широкое распространение спектральные методы исследования строения и идентификации органических соединений, что, в свою очередь, вызвало огромный рост потока информации по молекулярной спектроскопии. Это обстоятельство уже в настоящее время создает практически непреодолимые трудности при ручном поиске и обработке массивов спектральных данных и настоятельно требует использования для указанной цели ЭВМ (см., например, [1—5]). Однако объем памяти и быстрдействие современных ЭВМ пока еще недостаточно велики (см., например, [6]) для хранения и обработки детально закодированных спектров большого числа соединений. Кроме того, поиск путей сжатия спектральной информации без существенной ее потери (это позволило бы резко сократить как необходимый объем памяти, так и длительность решения каждой задачи на ЭВМ) в любом случае остается актуальной задачей. Одному из удобных методов для осуществления такого сжатия информации и посвящено данное сообщение.

Очевидно, каждый спектр (масс-спектр, ЯМР, ИК и др.) можно рассматривать как распределение некоторой случайной величины (случайную реализацию из генеральной совокупности), причем аргументом (x) обычно служит переменная, характеризующая качество (массовое число, частоту и т. д.), а функцией (y) — количество (интенсивность линий). Это распределение можно попытаться описать с помощью обобщенных характеристик, в качестве которых, как известно, используют моменты различных порядков [7]. В рассматриваемом нами случае наиболее важными, по-видимому, будут моменты первого порядка, представляющие собой математическое ожидание (или взвешенное среднее значение) соответствующей величины:

$$M(x) = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i}; \quad M(y) = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i}.$$

Погрешности величин $M(x)$ и $M(y)$ могут быть оценены аналогично ошибкам косвенных измерений (ср. [8]):

$$S_{M(x)}^2 \simeq S_x^2 + \left[\frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i - n \sum_{i=1}^n x_i y_i}{(\sum_{i=1}^n y_i)^2} \right]^2 S_y^2,$$

$$S_{M(y)}^2 \simeq S_y^2 + \left[\frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i - n \sum_{i=1}^n x_i y_i}{(\sum_{i=1}^n x_i)^2} \right]^2 S_x^2,$$

где $S_{M(x)}^2$ и $S_{M(y)}^2$, S_x^2 и S_y^2 — дисперсии величин соответственно $M(x)$ и $M(y)$, x и y . Предложенный метод был опробован на спектрах ПМР и масс-спектрах. Работа проводилась на ЭВМ «Наири».

Из коллекции, содержащей около 2000 ПМР спектров [9], случайным образом были отобраны 159 спектров и для них рассчитаны параметры $M(x)$ и $M(y)$. Затем из тех

же 2000 спектров было произвольно отобрано еще 15 спектров и по вычисленным для них величинам $M(x)$ и $M(y)$ проведен поиск в массиве из 159 спектров, причем было принято, что разброс значений $M(x)$ и $M(y)$ лежит в пределах $M(x) \pm S_{M(x)}$ и $M(y) \pm \pm 3S_{M(y)}$, $S_x \sim 0,1$ мд и $S_y \sim 0$ (интенсивности нормированы на целое число протонов). Максимальное число выдаваемых по запросу спектров не превышало при этом 3.

В случае масс-спектров в качестве характеристик использовались целочисленный молекулярный вес и параметр $M(x)$, рассчитанный по шести наиболее интенсивным (20% и выше) линиям. Из коллекции, содержащей порядка 5000 спектров [10], случайным образом были отобраны 159 спектров и затем еще 15. При поиске последних в массиве из 159 спектров по целочисленному молекулярному весу и $M(x)$ (принято, что разброс значений $M(x)$ лежит в пределах $M(x) \pm 3S_{M(x)}$, $S_x \sim 0$, $S_y \sim 3\%$ абс.) максимальная выдача спектров на запрос была равна четырем.

Таким образом, в случае использования как спектров ПМР, так и масс-спектров предлагаемый метод позволяет рассматривать только две величины (вместо 6—16) для каждого спектра, причем после выборки по этим двум величинам для последующего детального сопоставления остается не более 1/40 от первоначального объема коллекции.

В заключение отметим, что подобная система отбора может оказаться полезной также и при ручных поисках спектральной информации, например, в виде дополнительных ключей к каталогам спектров.

Авторы выражают благодарность В. А. Коптюгу и В. А. Бархашу за полезные обсуждения и помощь в работе.

ЛИТЕРАТУРА

1. Ю. П. Дробышев, Р. С. Нигматуллин, В. И. Лобанов, И. К. Коробейничева, В. С. Бочкарев, В. А. Коптюг. Использование ЭВМ для опознания химических соединений по спектральным характеристикам. (Система поисков ИК-спектров соединений).— Вестник АН СССР, 1970, вып. 8.
2. R. M. Silverstein, G. C. Bassler. Spectrometric Identification of Organic Compounds. New York, London, John Wiley & Sons, 1964.
3. F. Erni, J. T. Clerc. Anwendung von Computern bei der Strukturaufklärung organischer Verbindungen mit Hilfe spektroskopischer Methoden.— Chimia, 1970, v. 24, № 11.
4. L. H. Cross, J. Haw, D. J. Shields. Retrieval of infrared data.— In "Molecular Spectroscopy". London, Petroleum Institute, 1968.
5. Ю. П. Дробышев, Р. С. Нигматуллин, В. И. Лобанов, И. К. Коробейничева, В. С. Бочкарев, В. А. Коптюг. Машинная система идентификации хромофорных групп органических соединений по их электронным спектрам поглощения.— Семинар «Использование вычислительных машин в спектроскопии молекул». Тезисы докладов. Новосибирск, 1971.
6. W. Simon, J. T. Clerc. Moderne Methoden zur Struktur aufklärung Organischer Verbindungen.— Pure Appl. Chem., 1971, v. 26, № 1.
7. Е. С. Вентцель. Теория вероятностей. М., «Наука», 1969.
8. В. В. Налимов. Применение математической статистики при анализе вещества. М., «Наука», 1960.
9. W. Brügel. Nuclear Magnetic Resonance and Chemical Structure, v. 1. New York, London, Academic Press, 1967.
10. A. Cornu, R. Massot. Compilation of Mass Spectral Data. London, Heyden & Son Ltd. in Co-operation with Presses Universitaires de France, 1966.

Поступило в редакцию
11 января 1972 г.