

УДК 543.42

© Д. Б. Архипов, А. Л. Буляница, Л. В. Новиков

СЕДЬМОЙ ЗИМНИЙ СИМПОЗИУМ ПО ХЕМОМЕТРИКЕ (Санкт-Петербург, Роцино, 15–19 февраля 2010 г.)

В работе рассматривается тематика последнего Седьмого зимнего симпозиума по хемометрике — относительно новой для российского научного сообщества научной дисциплине. Дана оценка особенностям выбора тематики докладов, а именно, в какой мере рассматриваемые применения хемометрических методов в качестве математического обеспечения различных видов химического анализа традиционны или имеется определенная специфика формирования программы симпозиума, например по причине наличия каких-либо дополнительных требований к составу участников.

Кл. сл.: хемометрика, метод главных компонент, многомерный анализ, вейвлеты, наукометрия

Как сформулировано в Википедии: "Хемометрия или хемометрика (от англ. *chemistry* (химия) и *metrics* (как в "эконометрике" или "психометрии")) — раздел аналитической химии, ставящий целью получение химических данных с помощью математических методов обработки и добычи данных". Однако, как отмечено там же, однозначного определения, что же такое хемометрика, на сегодняшний момент нет и, вряд ли, когда появится.

Хемометрика — сравнительно новая научная дисциплина, находящаяся на стыке химии и математики, предметом исследований которой являются математические (преимущественно статистические) методы обработки данных, полученных в ходе физико-химических экспериментов.

Наиболее популярное определение принадлежит Д. Массарту [1]: "Хемометрика — это химическая дисциплина, применяющая математические, статистические и другие методы, основанные на формальной логике, для построения или отбора оптимальных методов измерения и планов эксперимента, а также для извлечения наиболее важной информации при анализе экспериментальных данных".

Несмотря на некоторые расхождения в формулировках следует признать, что хемометрический подход в первую очередь ориентирован на большие массивы данных и должен применяться для извлечения полезной информации. Суть этого подхода, по мнению автора [2], выражена в двух принципах. Во-первых, это понижения размерности задачи с помощью проекционных методов и, во-вторых, это использование формальных линейных моделей для объяснения связей в данных. Необходимо учитывать, что в русском языке до сих пор не сложилась общепризнанная система хемометрических терминов. Кроме того, число науч-

ных коллективов в России, теоретически развивающих хемометрический подход, до настоящего времени невелико (порядка 20).

Прошедший в 2010 г. симпозиум продолжил традиционную серию конференций в рамках международной программы, условно названной *Druzhbametrics*. Первые 3 мероприятия, названные Зимними школами-конференциями, были проведены в разных регионах России (в Костроме, на Алтае и в Пушкинских Горах) в 2002, 2003 и 2004 годах. Далее они получили название Зимних симпозиумов, в рамках которых также проводились школы. Последующие 4, 5 и 6-й симпозиумы были проведены в Черноголовке, Самаре и Казани в 2005, 2006 и 2008 годах соответственно.

На 7-м Зимнем симпозиуме по хемометрике "Современные методы обработки данных" было представлено 78 докладов. Официальным языком симпозиума был английский. В работе симпозиума участвовали 87 человек, из которых 47 представляли Россию, 6 — Прибалтику, 34 — дальнее зарубежье (США, Германия, Великобритания, Италия, Нидерланды, Польша, Венгрия, Австрия, Дания, Швеция, Норвегия, Финляндия, Ирландия, Австралия, Чили и ЮАР). На Российскую академию наук приходится 10 участников, на российские учебные заведения — 30. Российские ученые сделали 36 докладов, иностранные — 30. Авторами 12 докладов были как российские, так и зарубежные ученые.

До настоящего времени наиболее востребованным в хемометрике является метод главных компонент, основы которого предложены К. Пирсоном еще в 1901 г., лежащий в основе целого класса проекционных методов, используемых для сжатия исходных данных. Метод главных компонент особенно эффективен в задачах идентификации, дис-

криминации или классификации. Для задач, связанных с калибровкой, максимально востребован фундаментальный хемометрический метод PLS — метод частичных наименьших квадратов (Partial Least Squares), который благодаря созданию метода ортогональных PLS (OPLS) развивается в сторону задач, связанных с многовариантной калибровкой. В диссертации [2, с. 12] и статье [3] предложено трактовать аббревиатуру PLS как Projections on Latent Structures, что переводится как "проекции на латентные структуры". Всего в рамках симпозиума обсуждались 45 хемометрических методов. В задачах, связанных с распознаванием образов (pattern recognition) (по классификации [2], одного из трех основных разделов хемометрики), ни один из методов не является преобладающим. Формальным лидером (4 доклада) является вейвлет-преобразование, особенно эффективное в сочетании с другими методами. В работе [4], представленной совместно учеными Алтайского государственного университета и Университета Ольборга (Дания), сообщается о диагностике болезни Альцгеймера на ее ранней стадии методом магнитно-резонансной томографии. Для максимально четкого распознавания образов использовалось сочетание классических хемометрических методов, вейвлет-преобразования, текстурного анализа и метода опорных векторов (Support

Vector Machines (SVM)), разработанного для задач, учитывающих нелинейные спектральные зависимости [2].

Обращает внимание отсутствие доклада по тенденциям развития хемометрики, тогда как на конференциях сопоставимого уровня принято делать пленарные доклады, например по тенденциям аналитического приборостроения [5] или масс-спектрометрии [6]. Возможно, что ни в мировом, ни в российском сообществах хемометристов нет лидера, который бы владел всеми современными хемометрическими методами.

Исследование специфики формирования программы симпозиума было проведено на основе анализа распределения докладов по методам химического анализа, к которым были применены различные хемометрические процедуры. Исходные данные были получены наукометрическим методом по стандартной методике [7]. 70–75 % докладов симпозиума имели непосредственное отношение к аналитической химии и/или научному приборостроению. Полученные результаты были сопоставлены с последним обзором по хемометрике в *Analytical Chemistry* [8], где рассматривалось 145 работ, опубликованных в январе 2006 г.—декабре 2007 г. Тематические распределения докладов иллюстрируются данными таблицы.

Распределение численности публикаций по видам анализа

Вид анализа	VII Зимний симпозиум	Журнал <i>Analytical Chemistry</i>
ИК-спектроскопия	14	19
Рамановская	4	14
УФ-видимая область	10	11
Флуоресценция	3	6
Масс-спектрометрия (MS) в целом	13	17
В т.ч. GC-MS ^(*)	3	6
В т.ч. LC-MS ^(*)	3	5
Хроматография (без MS)	7	15
Электрофорез	4	4
Электрохимия в целом	6	4
В т.ч. сенсоры	6	2
В т.ч. спектро-ЭХ	0	1
ЯМР	1	4
Остальные	8	6
Всего	78	145

(*) Здесь GC и LC — газовая и жидкостная хроматография соответственно

По каждому виду химического анализа вычислялась доля от общего числа докладов. Соответствующие относительные частоты (рассчитанные по данным второго и третьего столбцов таблицы) сопоставлялись друг с другом. Степень их зависимости может быть оценена с помощью корреляционного анализа. Расчеты оценок коэффициента корреляции на основе 4 распространенных методик дают результаты: выборочный коэффициент корреляции (по Пирсону) 0.763, медианный аналог выборочного коэффициента корреляции (по Шевлякову) 0.675, ранговый коэффициент корреляции (по Спирмену) 0.648, знаковый (по Кендаллу) 0.495. Выбор коэффициента корреляции как меры зависимости не является единственно возможным. Однако оценивание близости распределений, например, на основе информационного расхождения Кульбака бесперспективно, поскольку имеется одно измерение (относительная частота), равное нулю. В этом случае, независимо от остальных данных, информационное расхождение будет бесконечно большим.

Приведенные оценки коэффициента корреляции в пределах от 0.495 до 0.763 допускают одинаковую интерпретацию наличия сильной корреляционной связи нелинейного характера. Тем самым можно полагать, что тематика симпозиума в достаточной степени отражает мировые тенденции развития хемометрических методов. При этом, однако, нет жестко заданных общепризнанных принципов отбора тематики докладов.

Как следует из данных таблицы, на симпозиуме наиболее широко были представлены ИК-спектроскопия в разных диапазонах (14 докладов) и УФ-видимая абсорбционная спектроскопия (10 докладов). Спектроскопии комбинационного рассеяния (рамановская спектроскопия) и флуоресцентному анализу посвящены 4 и 3 сообщения соответственно. 13 докладов посвящены масс-спектрометрии (МС), в том числе 3 — газовой хромато-масс-спектрометрии (ХМС) и 3 — жидкостной ХМС. По хроматографии без МС было сделано 7 докладов, по электрофорезу — 4 доклада. Основной областью применения жидкостной ХМС и всех видов электрофореза является анализ биополимеров. Обработкой соответствующих результатов занимается биоинформатика, которая теоретически является смежной областью с хемометрикой, но практически развивается отдельно. Более того, геномика, протеомика и другие "омики" используют для обработки результатов математические методы, имеющие между собой мало общего. В современной России наиболее востребованным направлением хемометрики для ХМС является анализ биологических жидкостей спортсменов на всевозможные запрещенные препараты. В Сочи специально создан исследовательский

центр РАН, одним из сотрудников которого является А.Л. Померанцев — сопредседатель рассматриваемого симпозиума [9]. В электрохимии хемометрика применяется реже — 6 докладов (все по сенсорам), в т. ч. 2 — по "электронным языкам". По 2 статьи приходится на акустические сенсоры, термический анализ, рентгеновский структурный анализ и радиохимические методы.

Сопоставление распространенности хемометрических методов в приложении к различным видам химического анализа с развитием собственно самих методов анализа, по-видимому, может позволить выявить специфику хемометрики. Однако, подобное перспективное исследование остается за рамками данной публикации.

СПИСОК ЛИТЕРАТУРЫ

1. *Massart D.L.* Chemometrics: A Textbook. N.Y.: Elsevier, 1988.
2. *Родионова О.Е.* Интервальный метод обработки результатов многоканальных экспериментов. Дис. ... док-ра физ.-мат. наук. СПб.: ИАП РАН, 2009. 272 с.
3. *Loo R.L., Coen M., Ebbels T. et al.* Metabolic Profiling and Population Screening of Analgetic Usage in Nuclear Magnetic Resonance Spectroscopy-Based Large-Scale Epidemiologic Studies // *Anal. Chem.* 2009. V. 81, N 3. P. 5119–5129.
4. *Dmitriev A., Kucheryavsky S.* Analysis of MRI Images of a Brain with Dementia Using Multivariate Approach // *Modern Methods of Data Analysis.* СПб., 2010. P. 46.
5. *Курочкин В.Е.* Тенденции развития аналитического приборостроения // Тез. III Всеросс. конф. "Аналитические приборы". СПб., 2008. С. 3.
6. *Галль Л.Н.* Итоги и перспективы развития масс-спектрометрии // Тез. III Всеросс. конф. "Аналитика России". Краснодар, 2009. С. 18–21.
7. *Архипов Д.Б.* II Съезд биофизиков России // *Научное приборостроение.* 1999. Т. 9, № 4. С. 114–116.
8. *Lavine B., Workman J.* Chemometrics // *Analytical Chemistry.* 2008. V. 80, N 12. P. 4510–4531.
9. *Rodionova O.Ye., Pomerantsev A.L., Houmoller L.P. et al.* Another Proof that Chemometrics Is Usable: NIR Confirmed by HPLC-DAD-MS and CE-UV // *Modern Methods of Data Analysis.* СПб., 2010. P. 32.

**Институт аналитического приборостроения РАН,
Санкт-Петербург**

Контакты: Буляница Антон Леонидович,
antbulyan@yandex.ru

Материал поступил в редакцию 2.04.2010.

**SEVENTH WINTER SYMPOSIUM ON CHEMOMETRICS
(Saint-Petersburg, Roshchino, February 15–19, 2010)**

D. B. Arkhipov, A. L. Bulianitsa, L. V. Novikov

Institute for Analytical Instrumentation RAS, Saint-Petersburg

The paper deals with the report themes of the last Seventh Winter Symposium on Chemometrics which is a relatively new discipline for the Russian scientific community. Estimation of peculiarities selection for themes of reports is given. The authors discuss if the studied chemometrical method application is traditional mathematical software or has definite specificity based on forming of symposium program because of additional requirements to the list of participants.

Keywords: chemometrics, principal components analysis, multivariate analysis, wavelets, scientometrics