

Методы, подходы и инструменты многофакторной хемоинформатики и биоинформатики в геоархеологических и археоминералогических исследованиях на спектроскопической и спектрографической базе

В археоминералогии и геоархеологии весьма широко применяются физические методы [Aitken, 1961] определения химизма и изотопии, основанные на использовании ряда спектральных или псевдоспектральных методов, являющихся потенциальными источниками хеометрической информации. В систематической монографии (являющейся одновременно ведущим руководством в химической археометрии) «Analytical Chemistry in Archaeology» [2007] рассматриваются, в частности, следующие аналитические методы спектральной археометрии: оптическая спектрофотометрия; инфракрасная абсорбционная спектроскопия / вибрационная спектроскопия; спектроскопия комбинационного рассеяния (т.е. рамановская спектроскопия); оптическая эмиссионная, атомно-абсорбционная и атомно-эмиссионная спектроскопия (в т.ч. – последняя с индуктивно-связанной плазмой); масс-спектрометрия с индуктивно-связанной плазмой / изотопная масс-спектрометрия; ряд методов радиационного анализа, включающих, помимо стандартных рентгенофлуоресцентных методов, подробно рассмотренных в книге [X-Ray, 2012], экзотические методы, такие как протонно-индуцируемая рентгеновская эмиссия (PIXE – proton induced X-ray emission), индуцируемая корпускулярной бомбардировкой эмиссия в γ – диапазоне (PIGME – Particle induced gamma ray emission), рентгеновскую фотоэлектронную спектроскопию и т.д.. Таким образом, основным объединяющим звеном в археометрических методах анализа является использование спектроподобных распределений [Злоказов, 1985], к которым, по определению, относятся формы одно- и многомерных спектров первичного или вторичного гамма-излучения, оптических спектров, спектров рентгеновского излучения, а также одно- и многомерных дифрактограмм, одно- и многомерных спектров распада, т.е. большинства тех методов, которые используются в археометрии, включая нейтронно-активационный анализ (в т.ч. – на базе эпитепловой нейтронной активации) и рентгеноструктурный /-дифракционный анализ. Следовательно, с метрологических позиций, исчерпывающее описание образцов того или иного археологического, археоминералогического, геоархеологического происхождения может быть произведено на базе единой программно-аппаратной системы обработки данных спектров и установления колокализации их в доверительных интервалах на разных образцах.

Таким образом, задача полноценной квантификации геоархеологических образцов и кластеризации их по свойствам в некотором многомерном пространстве признаков является, по существу, задачей хеометрики. Задачи совмещения хеометрики и археометрики ставят с конца 1980-х – начала 1990-х гг. [Hayek et al., 1991], однако односторонние пути их решения (например, на базе только масс-спектрометрических данных) препятствовали их решению до последних лет – до первых успешных попыток введения многофакторной археометрической классификации с использованием данных различных методов [Tomassetti et al., 2014]. Однако, отсутствие полной характеристики образца как химического, соответственно – гео-

химического / геоархеологического целого, не дает возможности сопоставлять результаты кластеризации, полученные на основе различных методов или их совокупностей различными группами. Для охарактеризования фоссилизации и биоминерализации в геологии / минералогии используются подходы геохимического генеза и методы анализа замещения с привязкой к тафономическим условиям среды. Вместе с тем, с точки зрения физической химии, происходит формирование пространства фазовых признаков – фазовых диаграмм (диаграмм состояний), сводимых к химико-термодинамическим основам – равновесным состояниям образца как физико-химической системы в условиях соответствия координатам фигуративной точки (что хорошо рассчитывается методом CALPHAD). Тот же принцип или подход может быть положен в основу методов интерпретации данных анализа в геоархеологии / археоминералогии, так как необходимо рассматривать объект-образец на его историческом и зонально-генетическом пространстве признаков, в размерности которого при наличии соответствующих данных может быть включена шкала времени, взятая из датировки образцов стандартными методами. Такой подход представлял бы также интерес как принцип анализа эволюции культур по прямым аналитическим данным. В самом грубом приближении для демонстрации можно проиллюстрировать это следующим примером: фазовые диаграммы и спектрально-колокализационные карты характеристических образцов энеолита по понятным причинам должны быть отличными от фазовых диаграмм и спектрально-колокализационных карт характеристических образцов палеолита, так как изменение уровня обработки материала и характера / состава / происхождения преимущественно используемых материалов приводит к кардинальному изменению статистики фазовых диаграмм в культуре, что в переходе между периодами или эпохами зачастую сопровождается полной элиминацией фазовых диаграмм устаревших материалов или материалов с утерянной технологией производства (точно также, как в археометаллографии, являющейся по методологии / метрологии частью металлографии, можно различить металлы и сплавы различных эпох по их структуре, отвечающей фазовым диаграммам их формирования, обусловленным существовавшей технологией).

Однако, CALPHAD и его аналоги исходят из принципов соответствия и непрерывности фаз, причем, согласно принципу соответствия, каждому комплексу фаз соответствует паттерн (геометрический образ) на диаграмме состав-свойство вещества образца. Данная диаграмма является графической репрезентацией корреляции между химизмом системы и значениями ее физико-химических характеристик – от электрохимических до спектральных. Таким образом, описание вещества образца на языке фазовых диаграмм является прямым коррелятом метода построения количественных моделей «структура-свойство» с использованием дескрипторов, взятых из аналитической практики или структурного анализа (QSPR – Quantitative Structure–Property Relationships). К таковым относятся дескрипторы, которые могут быть извлечены из спектроподобных распределений, в том числе – структурно-химического происхождения. К сожалению, в силу ограниченности списка литературы, мы не можем привести необходимые ссылки, что будет сделано в отдельной статье, но можем перечислить дескрипторы, которые могут быть репрезентативны для структурной археоминералогии. Так, структурное значение имеют дескрипторы, выводимые *ab initio* и находимые полуэмпирически: частичные заряды на атомах, частичные порядки связей, индексы активности / реакционной способности Фукуи (индексы свободной валентности, нуклеофильная и электрофильная суперделокализуемость), энергии катионной, анионной и радикальной локализации, энергии граничных молекулярных

орбиталей (ВЗМО и НСМО), дипольный и высшие мультипольные моменты распределения электростатического потенциала [Karelson *et al.*, 1996], а также дескрипторы молекулярных полей и, отчасти (с кинетической привязкой), константы заместителей. Все эти данные могут быть получены для любого минимального фрагмента-аналита образца геологического или минералогического, а значит – и геоархеологического / археоминералогического происхождения. Другие методы, в частности – методы электрохимии, часто используемые в археометрии и археоконсервации, в том числе сопряженные с электроимпедансной / кондуктансной спектроскопией, дающей при построении частотной зависимости спектроподобные распределения [Domenech-Carbo *et al.*, 2009], представляют также источники данных физико-химических дескрипторов QSPR. Дескрипторы же молекулярного подобия при применении молекулярно-кристаллического подхода к учету корреляционных поправок в теории химической связи могут говорить при «археометрическом скрининге» о подобии образцов или о культурной эволюции при построении «гомологических рядов» для множеств образцов (с учетом расположения в культурном слое и миграции в географическом пространстве) при анализе данных дескрипторов с применением методов ближайшего соседа (k-NN), опорных векторов (SVM) или РАМ (определение скрытой эвристической структуры в множествах уже кластеризованных статейных данных с суррогатными ключами).

Данный подход, заимствованный в хемоинформатике, применимый в археологической хеометрике, является необходимым эвристическим довеском в анализе данных археометрии геоархеологических / археоминералогических образцов. Автор статьи [Злоказов, 1983] писал, в частности: «автоматическая классификация или автоматическое распознавание образов формулируют свои задачи как задачи построения отображений классов заданных (а тем самым – и явно описанных) объектов классам явных дескрипторов, однако в современной практике анализа данных задача часто ставится с максимальной степенью неопределенности и некорректности: задана $f(x)$, требуется выявить ее информационную структуру и помочь ... экспериментатору не просто получить ответ на интересующие его вопросы, но и поставить эти вопросы». Для археоминералогии и геоархеологии важно правильно поставить вопрос: какой максимум информации мы можем извлечь из данного образца, исходя из того, какие данные могут быть использованы для построения его культурно-эволюционной линии и антропогеографической шкалы? После этого, в соответствии с методами оптимального планирования эксперимента, можно построить тот перечень спектроподобных распределений, использование которых при кластеризации данных в избранных целях (по взятым критериям), является целесообразным.

В частности, перечень дескрипторов и методов сбора спектроподобных распределений (ведущих к получению массивов этих дескрипторов) зависит от задач. Для анализа образцов биоматериала (кость и др.) требуется несколько отличный набор дескрипторов, чем для сбора информации о косных антропогенных объектах культуры. QSPR здесь соприкасается с QSAR (через воздействие условий среды на организм – источник ископаемого биоматериала), а ряд характеристик образца исходно биохимического происхождения могут быть получены лишь с применением цифрового аналитического инструментария системной биологии (в частности, с использованием нотации SBGN в реконструкции метаболических путей, палеодиет и пр.). В связи с этим необходимо совмещение техник СОВАС (Computer Based Analytical Chemistry), ведущих к получению спектроподобных распределений, QSPR и SBGN для биогенных, в т.ч. – частично минерализованных образцов. Ранее этот

принцип был внедрен нами для анализа современного биоматериала, включая продукты биоминерализации [Орехов, Градов, 2014]. Этот анализ, включая анализ химизма причин ранней некрологии древнего человека, постнекротических изменений в среде захоронения и сопутствующих геохимических процессов является возможным в свете совмещения в единой программе анализа распределений органической масс-спектрометрии, неорганической и изотопной масс-спектрометрии (соответственно, системной биологии как совокупности наук о биоорганических средах, микроэлементологии, неорганической химии культурных материалов косной и биокосной природы, и изотопологии / изотопомики данных абиогенных, антропогенных и биогенных сред), следовательно – в свете многомасштабного и взаимно-комплементарного анализа спектроподобных распределений. В разработанной нами системе эти процедуры производились на ЭВМ, автоматизировавшей сбор данных КАМАК-крейтов, подключенных к разным аналитическим приборам, но предлагаемый нами принцип более широк, чем доступное нам его воплощение, и может быть экстраполирован на разные методы анализа, типы оборудования и программные имплементации блок-схем алгоритмов.

Литература

1. Aitken M. J. *Physics and Archaeology*, New York: Interscience Publishers, 1961. 181 p.
2. Pollard M., Batt C., Stern B., Young S.M.M. *Analytical Chemistry in Archaeology*, Cambridge – New York: Cambridge University Press, 2007. 400 p.
3. X-Ray Fluorescence Spectrometry (XRF) in Geoarchaeology (Ed. by: M. S. Shackley), New York – Dordrecht – Heidelberg – London: Springer, 2011. 231 p
4. Злоказов В.Б. Математические методы анализа экспериментальных спектров и спектроподобных распределений // Физика элементарных частиц и атомного ядра, 1985. Том 16, Вып. 5. С. 1126–1163.
5. Hayek E.W.H., Krenmayr P., Lohninger H., Jordis U., Sauter F., Moche W. GC/MS and chemometrics in archaeometry // *Fresenius Journal of Analytical Chemistry*, 1991. Vol. 340, Issue 3. Pp. 153–156.
6. Tomassetti M., Marini F., Campanella L., Coppa A. Archaeometric classification of ancient human fossil bones, with particular attention to their carbonate content, using chemometrics, thermogravimetry and ICP emission // *Chem. Cent. Journ.*, 2014, DOI: 10.1186/1752-153X-8-26.
7. Karelson M., Lobanov V.S., Katritzky A.R. Quantum-Chemical Descriptors in QSAR / QSPR Studies // *Chem. Rev.*, 1996. Vol. 96, Issue 3. Pp. 1027–1044.
8. Domenech-Carbo A., Domenech-Carbo M., Costa V. *Electrochemical Methods in Archaeometry Conservation and Restoration*, Berlin, Heidelberg: Springer, 2009. 166 p.
9. Орехов Ф.К., Градов О.В. Гибридизация COBAC, QSPR / QSAR и SBGN: единство теории и практики в анализе данных и проектировании спектрально-биохимического лабораторно-диагностического и биомедицинского оборудования. *Биотехносфера*, 3(33), 2014. С. 29–31.