

© В.А. Нагорский, Б.А. Горлицкий, В.В. Скворцов, 2008

УДК: 550. 8: 681.327.12

*Институт геохимии окружающей среды
НАН и МНС Украины, г. Киев*

РАСПОЗНАВАНИЕ ОБРАЗОВ В ГЕОЛОГИИ – ПРОБЛЕМЫ ПРИМЕНЕНИЯ СЕГОДНЯ

Распознавание образов (РО) – область прикладной математики, которой присущ кибернетический подход к решению разнообразных задач, а именно представление природных объектов, относимых в определенном отношении к разным типам (классам), в качестве информационных систем, и использование свойств последних как типических образов реальных объектов. Методы РО широко применяются для решения различных задач как прикладной, так и теоретической геологии. Пик популярности методов РО приходится на 1970–1980-е годы, когда редкая публикация в геологических изданиях на тему “математизации”, “автоматизации” или “компьютеризации” обходилась без упоминания об РО. Однако позднее эти методы как бы “ушли в тень”; возможно, их действительно стали использовать относительно реже, чем раньше (сказалось некоторое разочарование на фоне завышенного ожидания), но скорее – это следствие общего спада объёмов научных исследований в стране с начала 1990-х годов и коммерциализации науки с ее “секретами производства”. Особую роль, на наш взгляд, сыграла в этом, как ни странно, возросшая доступность методов РО, легкость их использования в условиях, когда большинство различных современных пакетов обработки и анализа данных оснащены и “подпакетом” программ РО.

Массовая доступность современной компьютерной техники и соответствующих технологий (включая геоинформационные) породила опасную, по мнению авторов, тенденцию. Если заострить проблему, то это – “легкость нажатия кнопок”, вместо изучения трудных основ прикладной математики. Давно известно, что роль творческого элемента и сложность освоения возрастают в последовательности hardware – software – brainware (т. е. “железо” – программы – “мозги” (алгоритмы)). Сегодня “эффект Митрофанушки” – зачем географию изучать, если всюду есть извозчики – уже весьма явно ощутим и во многих работах по геоинформатике (в частности, с использованием РО) проглядывает не просто нежелание автора, так сказать, углубляться в теорию, но и его искреннее убеждение в том, что это излишне. Доходит до курьезов: в одной из работ метод (способ) циклического поточечного контроля качества распознавания “джекнайф”, т. е. “складной

нож”, был назван “методом Джекнайфа”. Внешне – явная мелочь, однако здесь просматривается “верхушка айсберга” того массового снижения потребности изучения теоретических основ обработки информации, о котором шла речь выше. Распознавательский подход весьма неоднозначен по своей сути. Вместе с тем известно много конкретных методов (алгоритмов) РО и много формальных типов геологических задач (ситуаций), в которых, в принципе, можно применить РО. Однако не только эффективность (качество) решения каждой реальной задачи зависит от того, какой конкретно алгоритм мы выберем (и как выберем?); большинство комбинаций “метод – задача” либо приводят к неустойчивым ненадежным решениям, либо могут дать решение в виде “чепухи”. И многие сегодняшние исследователи (особенно молодые) отдают предпочтение “нажатию кнопок”, даже не подозревая о тех “подводных камнях”, которые их ожидают.

Авторы попытались выделить явно неоднозначные (сложные, спорные) аспекты применения методов РО в геологии, особенно те, где существуют уже достаточно устойчивые заблуждения. Это своего рода мифы о распознавании образов.

Научные мифы – это устойчивые мнения, которые верны лишь отчасти, верны лишь при определенных условиях либо вообще не верны. В данном контексте они относятся к “универсализму” и высокой “интеллектуальности” методов РО, их эффективности, соотношению локальных и глобальных методов, возможности повысить надежность распознавания путем усреднения результатов многих методов РО и др. В целом авторы призывают использовать эти методы с большой осторожностью и осмотрительностью.

Ниже предлагается и обосновывается следующая схема действий в “распознавательских ситуациях”.

1. Изучить особенности конкретных алгоритмов РО до того, как пытаться их применять на практике.
2. Постараться изменить постановку исходной задачи так, чтобы она перешла в другую, более надежно решаемую формальную схему. Например, вместо распознавания мелких и средних месторождений нефти и газа, введя в исходные данные (“эталонные выборки”) запасы углеводородов этих объектов в явном виде, заменить схему задачи на регрессионную, т. е. оценивать ожидаемые запасы будущих (прогнозных) объектов, после чего их разделение на группы уже не составит труда.
3. Если такой замены сделать не удастся, то использовать строгий математический (а не эвристический – как в распознавании образов) метод дискриминантного анализа, желательно в его наиболее общей квадратичной, форме. Для этого следует предварительно провести максимально возможную нормализацию обеих эталонных выборок данных.

4. Если и этого сделать не удастся (малые объемы выборок, их очень высокая вариабельность, внутренняя неоднородность и др.), то использовать методы визуализации внутренней структуры многомерных выборок (факторный анализ и др.) для предварительной экспертной оценки того, какие именно типы алгоритмов РО могут быть наиболее успешными и эффективными в данных конкретных условиях.
5. Применить не один, а несколько близких по свойствам алгоритмов РО, отвечающих выводам предыдущего пункта, однако их результаты не “усреднять”. Методы рекомендуется применять в форме “джекнайфа”, как наиболее объективной формы контроля качества распознавания; выбирается один метод, показавший лучший результат.