**Некоторые проблемы формализации гуманитарных знаний (на примере археологии)**

Ф.П. Тарасенко, Томский Государственный Университет

**1. Вступление**

В свое время еще И. Кант сформулировал мысль, которую впоследствии повторяли и интерпретировали многие философы, о том, что любая отрасль знания с тем большим основанием может называться наукой, чем чаще и успешнее она использует математику в собственных целях. С позиций современной системологии, любая дисциплина в процессе развития проходит путь от первых, простейших, вербальных описаний предмета своих интересов, через все большую детализацию, уточнение, специализацию собственного профессионального языка, до того состояния, когда становится возможным описывать изучаемые явления и процессы не приблизительно и расплывчато, а все более конкретно и точно, используя для этого языки все более абстрактных разделов математики. Каждый шаг вперед на этом пути есть результат приобретения новой информации об изучаемом объекте: сами по себе математические методы не дают новой информации; они лишь позволяют преобразовать имеющуюся в постановке задачи информацию в такой вид, который более обозрим, удобен, полезен, пригоден для наших целей.

Такое представление о прогрессе знаний о мире подтверждается на примерах различных наук. И если какая-то из них слабо или вообще не пользуется математикой, то это, говорит лишь о большой сложности, т.е. малой доступности информации, слабой изученности предмета исследований, о том, что в этой дисциплине еще много неизведанного и у нее большие перспективы развития.

Важным событием для любой отрасли науки является ее успех в использовании математических методов; в особенности отрадно видеть это в гуманитарных науках. Интересным примером служит изданная в Новосибирске монография "Математические методы в археологических реконструкциях" под ред. А.П. Деревянко и Ю.П. Холюшкина" (СО РАН, 1995г.). Не будучи археологом, не берусь судить о содержательной стороне и значимости рассматриваемых в этой книге конкретных задах. Но, возможно, коллегам-археологам могут оказаться полезными некоторые соображения, возникшие у прикладного математика и статистика после прочтения их книги. Тем более, что приходилось встречаться с очень похожими ситуациями при приложении математики в других науках - геологии, биологии, психологии, социологии. Подчеркну, что общее впечатление от книги уважаемых археологов весьма положительное, несмотря на то, что последующие замечания и комментарии носят несколько критический характер.

**Системность мира и ее разные описания**

Рискну утверждать, что подлинная научность знаний проявляется не тогда, когда мы переходим к адекватным математическим моделям (это уже завершающий этап развития теории), а гораздо раньше - при осознании системности объекта исследования, т.е. при учете не только специфических особенностей объекта, но и его целостности, его неотъемлемости от окружающей среды. Замечательно, что для авторов монографии это уже пройденный этап.

Однако, знания о любых (в том числе и изучаемых археологией) системах могут выражаться в разных формах. Известны такие вариации системологии, как тектология, общая теория систем, прикладной системный анализ, объектно-ориентированный подход, системотехника, исследование операций, синергетика, и ряд других. Авторы книги отдают явное предпочтение объектно-ориентированному подходу (ООП). Конечно, в каждом конкретном случае необходимо придерживаться какой-то конкретной методики. Но объявлять любую методику "наиболее полной и комплексной", "наиболее универсальной и эффективной" (см. стр. 5), "наиболее общей" (стр. 11) - (уместно да и то не всегда) в эмоциональных (политических, религиозных, межличностных) сферах, но никак не научных. У каждой из методик есть свои достоинства и недостатки; например, технология "мягких" систем в прикладном системном анализе мне представляется более удобной при исследовании слабо и плохо структурированных объектов, нежели ООП, предназначенный для программирования ЭВМ. А ведь авторы книги сами подчеркивают размытость многих археологических реконструкций (стр. 22, 122).

**Классификация как простейшая модель разнообразия**

Различное понимание одних и тех же слов часто создает проблемы - не только в быту при употреблении разговорного языка, но и в науке при пользовании профессиональными языками. В этом причина большого значения, придаваемого терминологии учеными. В этой связи обращает на себя внимание очень расширительное толкование авторами книги термина "классификация". Под этим словом они понимают и Периодическую систему элементов, и периодизацию истории по общественно-экономическим формациям (стр. 12.), и разработанную ими семантическую карту классификатора (стр. 13 и многократно далее). Между тем, существует возможность избавить данный термин от смысловой перегрузки.

Если согласиться с тем, что все наши знания существуют в форме моделей, то реальному миру соответствует мир отображающих его моделей. Многообразие моделей, их многокомпонентность, вложенность друг в друга (от исходных понятий до развитых теорий) ставит вопрос о том, что же является "атомом" в мире моделей. Для абстрактных моделей элементом, т.е. простейшей моделью, на базе которой строятся все остальные модели, является классификация.

Классификация позволяет в самом упрощенном, конечном виде отобразить бесконечное разнообразие реального мира, в котором нет тождественных явлений, объектов, а тем более - субъектов. Достигается это путем введения понятий "сильных" и "слабых" отличий, объединения "слабо" отличающихся объектов в одну группу, и отказа вообще различать "слабые" отличия внутри этой группы, называемой классом. Понятно, что число классов и границы между ними субъективны, т.е. определяются нашим пониманием оценочных слов "сильная" и "слабая" для разницы между рассматриваемыми объектами. Классификация может быть субъективной ("произвольная" классификация, основанная на целях субъекта, строящего ее) и менее субъективной ("естественная" классификация, учитывающая природную кластеризацию в рассматриваемом множестве). Но в любом случае классификация - лишь модель, приблизительно описывающая реальность. Более сложные модели появляются за счет учета все более мелких различий (например, иерархическая классификация типа предложенной авторами схемы), либо за счет рассмотрения новых, дополнительных отношений между классами (например, порядковая шкала, где учитываются не просто отличия между классами, как в номинальной шкале, но и отношения предпочтения между ними; другой пример этого - таблица Менделеева), либо за счет введения "перекрывающихся" классов (например, статистические и расплывчатые классификации).

В свете сказанного хотелось бы прокомментировать разработанный авторами "стандартный вариант классификационного фрагмента" (стр. 13), используемый далее везде по всей книге. Основанный на фундаментальных понятиях диалектики, он, несомненно, обладает большой общностью, чем авторы заслуженно гордятся. И все же объявлять его "наиболее общим" и единственно правильным - явное преувеличение: как и всякая модель, любая классификация обладает пределами своей адекватности. Попытки же подогнать любой объект под единую классификацию на чем-то обязательно потерпят неудачу. Не избежали этого и авторы: если рис. 3-8 наглядно иллюстрируют полезность вводимой классификации, то рис. 20-25 не менее наглядно демонстрируют явную натянутость и наивность ее применения в других случаях (скажем, что делать, если ВАК предложит изменить количество оппонентов или у человека несколько высших образований?).

Ограниченность предложенной классификации заложена и в том, что она предусматривает соблюдение отношения эквивалентности на каждом уровне иерархии. В реальности это условие может не выполняться.

**Об измерительных шкалах и методах обработки данных**

Специального обсуждения заслуживает вопрос об измерительных шкалах: не только археологи недооценивают значение этого вопроса. Дело в том, что каждая шкала порождает данные, над которыми можно выполнять лишь допустимые для данной шкалы преобразования. Выполнение недопустимых операций приводит к абсурдным результатам. Многие удивлялись, почему в

80-х годах "не сработал" средний школьный балл при зачислении в вуз; а ведь для порядковой шкалы, к которой относятся и школьные отметки, операция сложения является недопустимой. Или, если вчера было 5° С, а сегодня 10° С, вы скажете, что сегодня "в два раза теплее", но что вы скажете, если вчера было 0° С, а сегодня 5° С? Или вчера -5° С, а сегодня +5° С? Дело в том, что для интервальной шкалы (в том числе - температурной) операции деления и умножения недопустимы.

Поэтому довольно частое различение только качественных и количественных признаков (что принято и в обсуждаемой книге) является недостаточным, как только приходится иметь дело с обработкой экспериментальных данных. А в археологии этого не избежать.

Различают две качественных шкалы (номинальную и порядковую) и четыре количественных (интервалов, циклическую, отношений и абсолютную). Кроме этих "типовых" шкал есть и производные от них - суперпорядка, нелинейные, дискретизованные, зашумленные и т.д. При обработке экспериментальных данных важно тщательно следить за тем, чтобы над ними выполнялись только допустимые операции, а они специфичны для каждой шкалы.

Второй важный момент обработки данных состоит в строгом различении прямых и косвенных измерений. Если то, что нас интересует, недоступно непосредственному наблюдению, но можно измерять некоторую величину, связанную с ним, то эти два признака могут принадлежать разным шкалам (это и будет косвенное измерение). И если наблюдаемая величина измеряется в более сильной шкале, при ее преобразованиях можно применять только операции, дозволенные в шкале исследуемого ненаблюдаемого прямо признака. Безусловно, такие ситуации бывают не только в биологии и психологии, но и в археологии.

Связанность ограничений на обработку данных с типом измерительной шкалы заставляет обратить внимание еще на одно широко (к сожалению) распространенное заблуждение о безоговорочной допустимости оцифровки качественных признаков. Увы, авторы книги разделяют это заблуждение(см. § 7.3.2. на стр. 105). Метод оцифровки сводится к "разумному" присвоению числовых меток категориям качественных признаков и их последующей обработке как чисел. Такой подход игнорирует проблему допустимых операций и может приводить к неверной интерпретации результатов.

Причин, по которым оцифровка многим кажется безобидной операцией, предоставляющей удобства при анализе данных, две. Во-первых, в некоторых алгоритмах анализа могут отсутствовать недопустимые операции, и тогда оцифровка действительно безопасна. Во-вторых, иногда слабые шкалы являются не настоящими качественными, а огрубленными количественными. При этом "испорченное" количество все-таки несет в себе количественное начало, для которого арифметические преобразования имеют какой-то смысл. И все же для номинальных и многих ранговых данных оцифровка в принципе недопустима. Тем более, что существуют алгоритмы совместного анализа разнотипных данных, т.е. измерений, принадлежащих различным шкалам.

**Об использовании статистических методов**

Книга заслуженное внимание уделяет пропаганде применения в археологии различных статистических методов (§ 2.2.2) и иллюстрирует их полезность на многих примерах (гл. VIII). Однако пропаганда статистики должна также сопровождаться предупреждениями о трудностях, ловушках и возможных неправильностях в использовании статистических методов. Пренебрежение ими дает основания для мрачно-юмористических высказываний типа "Есть три вида лжи - просто ложь, наглая ложь и статистика". Оставив в стороне человеческий фактор (например, недобросовестный или субъективный отбор данных), можно выделить несколько правил "статистической техники безопасности", связанных с самой природой статистических выводов.

Данные должны иметь действительно случайную природу (обладать статистической устойчивостью), что далеко не всегда имеет место, и очень непросто проверяемо.

Закономерность, выявленная статистически, никогда не бывает абсолютно точной: числовая характеристика всегда оценивается лишь приближенно; вероятность ошибки статистического вывода всегда отлична от нуля (ошибки первого и второго родов).

В структуре любого алгоритма статистической обработки данных заложены априорные предположения о природе данных. Если эти предположения расходятся с тем, что есть на самом деле, выводы получаются совсем не такого качества, которое от них ожидается. Практики же редко проверяют выполнимость "паспортных" ограничений на применимость процедур.

Выявленная статистическая закономерность подлежит содержательной интерпретации. Статистику нельзя обвинять за то, что вполне надежный статистический вывод неверно проинтерпретирован специалистом - пользователем; последний же часто "пеняет на зеркало".

В силу указанных причин очень верным является совет авторов археологам повышать свою квалификацию в области статистики (стр. 25), хотелось бы добавить к этому рекомендацию в сложных случаях почаще привлекать к сотрудничеству профессиональных статистиков. Например, для меня осталось загадкой, почему авторы проигнорировали (см. стр. 123 и гл. VII) результаты своих соседей из Института математики СО РАН профессоров Загоруйко и Лбова, которых они, очевидно, лично знают.

**Заключение**

Оценка монографии "Математические методы в археологических реконструкциях" (Новосибирск, 1995) археологами и историками - дело специалистов в этой области. Может быть, их заинтересуют и впечатления специалиста совсем из другой области - анализа данных, - изложенные в данной статье. Еще раз подчеркну: хотя здесь я изложил комментарии критического порядка, в целом обсуждаемая книга является несомненным шагом вперед в прогрессе археологии. Адекватное использование математики и компьютерных технологий есть явный признак повышения информационной зрелости любой науки, и за археологию можно порадоваться.