УДК 316:30:004

ПОКРОВСКИЙ А.М.

Применение новых информационных технологий в общественных науках: состояние и перспективы[1]

Анализируются возможности, предоставляемые современными информационными технологиями для математического моделирования социальных явлений и процессов по данным социологических исследований. Обосновывается продуктивность применения систем анализа данных к решению социальных проблем.

Ключевые слова: информационная технология, социальные явления, социальные процессы, социологические исследования, математическое моделирование, система анализа данных.

Проблемы применения методологии количественного моделирования в общественных науках волнуют представителей различных теоретических направлений, и в этой связи интересной представляется коллективная монография «Data Mining в эмпирической социологии: методология количественного моделирования», авторами которой являются представители трех наук: математики – профессор Шуметов В.Г., социологии – доцент Лясковская О.В. и психологии – докторант Гудова Т.Г. Рецензируемая совместная работа авторов, ведущих исследования в смежных областях наук, представляет особый интерес, поскольку до настоящего времени не появлялись работы, в которых столь детально рассматриваются возможности, предоставляемые использованием математического моделирования и новых информационных технологий в общественных науках вообще и в эмпирической социологии в особенности.

Примечательно, что рецензентами данной монографии выступили также представители разных областей наук – доктор философских наук, профессор Д.Г. Горин, заведующий кафедрой социально-гуманитарных дисциплин Брянского филиала Орловской региональной академии государственной службы, с одной стороны, и доцент кафедры алгебры и математических методов в экономике Орловского государственного университета, кандидат экономических наук С.П. Строев – с другой. Издательство ОРАГС – ныне Орловского филиала Академии народного хозяйства и государственной службы при Президенте РФ – любезно предоставило нам тексты рецензий профессора Д.Г. Горина и доцента С.П. Строева, что позволяет сопоставить их суждения, во-первых, с подходом авторов монографии к проблеме методологии количественного моделирования в общественных науках и, во-вторых, с мнением автора данной рецензии.

Прежде всего следует согласиться с констатацией профессором Д.Г. Гориным того факта, что до настоящего времени среди отечественных социологов наблюдается недостаточно глубокое использование результатов выполненных ими массовых социологических опросов. Одной из причин такой ситуации, на его взгляд, является недостаточное владение статистическими методами анализа данных вообще и многомерных методов в особен-

ности. Он отмечает, что в эмпирических социологических исследованиях социологи зачастую ограничиваются анализом «линейных» распределений вариантов ответов респондентов, и значительно реже встречаются попытки выполнить анализ так называемых парных распределений (таблиц сопряженности признаков). Тем самым результаты массовых опросов, сопряженных со значительными затратами времени и средств, используются явно недостаточно.

Далее профессор Д.Г. Горин отмечает, что широкое распространение такой информационной технологии, как система анализа данных общественных наук SPSS Base, которая является стандартом дефакто в анализе результатов социологических опросов, дает в руки социологу инструмент, позволяющий производить не только простейшие статистические процедуры типа вычисления распределения частот и построения таблиц сопряженности с целью выявления взаимосвязей между признаками, но и осуществлять их глубокую статистическую обработку – то, что получило название «Data Mining», в русскоязычной литературе – «добыча знаний из данных, интеллектуальный анализ данных». Он подчеркивает, что научиться использовать этот инструмент – это значит на порядок повысить эффективность использования дорогостоящих и трудоемких результатов социологических опросов.

Здесь уместно сказать несколько слов о применяемом авторами монографии инструментарии – системе анализа данных общественных наук SPSS Base (аббревиатура англ. «Statistical Package for the Social Sciences»), представляющей собой компьютерную программу для статистической обработки данных. Эта система развивается начиная с 1975 года, и в 2009 году компания SPSS произвела ребрендинг своего статистического пакета, который стал называться PASW Statistics (Predictive Analytics SoftWare). 29 июля 2009 года компания SPSS объявила о том, что она приобретается фирмой IBM, и в августе 2011 года появилась версия IBM SPSS Statistics 20. В России этот программный продукт поддерживается фирмой СПСС «Русь» (SPSS Rus), которая проводит более пятнадцати учебных курсов – от вводных, предназначенных для начинающих аналитиков, до продвинутых, рассчитанных на тех, кто уже имеет большой опыт в сфере анализа данных. Цены на эти учебные курсы достаточно высокие – от 3500–7500 руб. на дистанционные курсы до 15 000–60 000 руб. на регулярные курсы для незарегистрированных пользователей и 7500–30 000 руб. для зарегистрированных пользователей.

Помимо систем ***IBM*** SPSS Statistics на мировом рынке предлагаются программные продукты класса «Data Mining», предназначенные для пользователей, не имеющих специализированного образования в области статистической обработки данных и математического моделирования. Но создание таких программных продуктов связано с их удорожанием, дорогим является и обучение их применению. Так, по данным фирмы СПСС «Русь», стоимость двухдневного обучения незарегистрированных пользователей пакета SPSS Clementine v.12 – наиболее часто используемого инструмента Data Mining – составляет от 30 до 52 тыс. руб., т.е. использование инструментов Data Mining обходится весьма недешево.

Авторы рецензируемой монографии, исходя из своего практического опыта, пошли по другому пути. Они предлагают пользователю, а конкретнее – специалистам-социологам, применять для сложного анализа данных не дорогостоящие программные продукты, а сравнительно недорогую – порядка 1000 долл. – систему анализа данных общественных наук SPSS Base v.13.0. Как отмечает в своей рецензии доцент кафедры алгебры и математических методов в экономике Орловского государственного университета С.П. Строев, практический опыт авторов монографии показывает, что при наличии в учебном учреждении или научно-исследовательской организации консультантов, специализирующихся на применении разнообразных статистических процедур, вполне возможна их совместная работа с аналитиками-социологами, в результате которой и происходит «получение знаний

из данных». А необходимость в такой работе очевидна: отсутствие должного понимания возможностей технологий Data Mining приводит к тому, что основная масса результатов социологических опросов остается «за бортом». Пользователь, не искушенный в возможностях технологий Data Mining, ограничивается лишь анализом «процентов» распределения вариантов ответов респондентов, но это лишь начало аналитической деятельности по «извлечению знаний из данных».

Несколько слов об авторах рецензируемой монографии. Двое из них – доктор экономических наук, профессор В.Г. Шуметов и кандидат социологических наук, доцент О.В. Лясковская – являются сотрудниками кафедры математики и математических методов в управлении Орловского филиала РАНХиГС, причем они не новички в эмпирической социологии. Ими опубликованы монографии, посвященные проблемам моделирования и прогнозирования электорального поведения россиян, социального самочувствия и уровня жизни населения, включая проблемы оценки здоровья на региональном уровне. Кандидат психологических наук Т.Г. Гудова является докторантом кафедры социологии и психологии этой же академии и также не новичок в науке. Ее перу принадлежит монография «Психологическая работа при дезадаптационных расстройствах личности», подготовленная в соавторстве с О.В. Ефременковой и В.А. Чвякиным и опубликованная одним из центральных издательств в 2007 году, ряд статей в изданиях по списку ВАК. В рецензируемой монографии ее перу принадлежит глава, посвященная применению технологий Data Mining для решения проблем социально-психологической адаптации современного студенчества.

Учитывая адресность монографии, представляется достаточно логичной ее структура. Вначале авторы излагают методические аспекты технологии Data Mining в эмпирической социологии. Значительный интерес здесь представляют авторское видение современного состояния применения количественных методов анализа данных социологических исследований, метрологическая характеристика социологических данных, получаемых в результате массовых опросов респондентов и сбора экспертной информации, обзор методов их одномерного и многомерного анализа, а также характеристика информационных технологий интеллектуального анализа данных.

Затем авторы переходят к вопросам подготовки количественных данных по результатам социологических опросов, а также приводят описание одномерных процедур анализа, ограничившись главным – сравнением средних для различных социальных групп респондентов.

От одномерных методов анализа авторы переходят к многомерным методам, среди которых модели многомерного дисперсионного анализа, а также их частный случай – общая линейная модель. Рассматриваются и более сложные модели множественного линейного регрессионного анализа, в которых количественными являются не только результативные переменные (признаки), но и факторные, статусные.

Далее в монографии рассматриваются такие методы многомерного статистического анализа, как факторный анализ, являющийся эффективным методом выявления структуры переменных, кластерный анализ, который является средством многомерной классификации объектов. Далее рассматривается дискриминантный анализ как средство классификации «с учителем». Завершает описание методов анализа данных раздел, посвященный применению в эмпирической социологии нейросетевых нелинейных моделей.

Примечательно, что авторы не ограничиваются описанием возможностей тех или иных методов анализа социологических данных, но и демонстрируют их на конкретных примерах из практики своей аналитической деятельности. Тем самым образуется «мостик» между теорией статистических процедур и практикой их применения.

Однако рецензируемая монография – это не сборник «примеров», она имеет явно выраженную научную новизну. Если говорить кратко, то научная новизна монографического исследования состоит в разработке целостной технологии извлечения знаний из данных, содержащих ряд связанных меж-

ду собой этапов. Это прежде всего разработанный авторами метод подготовки количественных переменных по данным социологических опросов, а также методика количественного шкалирования порядковых переменных. Особенно важно, что предложенные авторами методы позволяют аналитику перейти от порядковых шкал к количественным с тем, чтобы затем применить и методы одномерной статистики, и такие мощные методы моделирования и прогнозирования, как множественный регрессионный и дискриминантный анализы.

Монография не лишена недостатков. На наш взгляд, недостаточное внимание авторы уделяют проблеме проверки валидности выборок. Нет достаточно четких указаний на требования к переменным, а они, например, в корреляционно-регрессионном и кластерном анализе несколько отличаются, недостаточное внимание уделено непараметрическим методам статистики. Представляется недостаточно обоснованным предлагаемый авторами метод количественного шкалирования порядковых переменных в случае более чем трех уровней соответствующих признаков, в монографии слабо представлены экспертно-аналитические модели, применяемые в условиях неопределенности информации.

Следует пожелать авторам устранить эти недостатки при переиздании монографии, которая, конечно же, послужит хорошим пособием не только для аналитиков-социологов, но и для студентов, изучающих методы анализа эмпирических данных, полученных в результате социологических опросов, а также статистических и эмпирических данных.

Подводя итоги, можно заключить, что монография «Data Mining в эмпирической социологии: методология количественного моделирования» представляет значительный интерес как для широкого круга социологов, использующих эмпирические методы исследования, так и для специалистов в других областях наук и может претендовать на один из лучших фундаментальных трудов по проблеме использования современных информационных технологий для глубокого анализа социальных явлений и процессов.

[1] Рецензия на монографию Шуметова В.Г., Лясковской О.В., Гудовой Т.Г. «Data Mining в эмпирической социологии: методология количественного моделирования» / Под общей редакцией профессора В.Г. Шуметова. Орел: ОРАГС, 2011. 275 с.