

О принципах формализованного качественного анализа социологических данных

М.А. Михеенкова

Аннотация. В работе характеризуется класс задач формализованного качественного анализа социологических данных, описывается структура интеллектуальных систем, предназначенных для решения этих задач. Обсуждается адекватность средств ДСМ-метода автоматического порождения гипотез исследуемой социальной реальности, рассматривается использование этих средств для решения задачи изучения детерминант поведения и его предсказания. Предлагаются инструменты для изучения рациональности мнений.

Ключевые слова: формализованный качественный анализ, интеллектуальные системы, автоматическое порождение гипотез, рациональность мнений.

Введение¹

В сложившейся практике социологических исследований словосочетание «формализованный качественный анализ» воспринимается как оксюморон. Использование формального аппарата считается несомненной прерогативой количественного – как правило, статистического, анализа. Качественные же стратегии анализа воспринимаются почти единственно как творческие эвристики общения исследователя с респондентом с последующим обобщением полученной информации и типологизацией эмпирического материала [1, 2]. Таким образом, как сбор данных, так и последующий анализ отдаются на откуп интуиции, опыту, наконец, способностям эксперта-социолога. Ощущение неразрывности обоих этапов социологического исследования, по-видимому, возникает именно из-за творческого характера получения исходных данных: решение задачи формального ана-

лиза результатов нарративного интервью или встроенного наблюдения действительно представляется нереализуемым при существующих возможностях формального анализа текстов.

Однако структура качественного анализа социологических данных, как она понимается, к примеру, в одном из наиболее развитых и авторитетных методов такого анализа – *обоснованной теории* (grounded theory) [3], – вполне воспроизводима формальными средствами. В самом общем виде *обоснованная теория* есть построение теории на основе эмпирических фактов с использованием индуктивного анализа, причем сбор информации происходит вплоть до насыщения выборки – когда новые индикаторы, категории и т.п. (термины *обоснованной теории*) перестают формироваться². Именно задача автоматического извлечения интерпретируемых зависимостей между раз-

¹ Работа выполнена при поддержке Президиума РАН (Программа фундаментальных исследований, 2009г.).

² Отметим предварительно, что идея индуктивного анализа именно насыщенной выборки была независимо реализована в одном из рассматриваемых ниже подходов – ДСМ-методе автоматического порождения гипотез (в его варианте для анализа социологических данных).

личными факторами, неявно содержащимися в массивах данных, является центральной для современных методов интеллектуального анализа данных (*knowledge discovery* [4])³.

Собственно для решения подобного рода задачи – извлечения интерпретируемых зависимостей из эмпирических данных с учетом особенностей предметной области – и развиваются методы формализованного качественного анализа социологических данных (ФКАСД). В западной практике широко представлен так называемый качественный сравнительный анализ QCA [5, 6], использующий для указанных целей аппарат булевой алгебры и нечетких множеств [7]. В отечественных исследованиях эта роль отводится ДСМ-методу автоматического порождения гипотез⁴ (ДСМ-метод АПГ) в его варианте для анализа социологических данных [8]. ДСМ-метод представляет собой специальный класс рассуждений (ДСМ-рассуждения), реализующий синтез познавательных процедур – эмпирической индукции, структурной аналогии и абдукции как средства принятия гипотез на достаточном основании [9]. ДСМ-рассуждения формализованы посредством бесконечнозначных логик степеней правдоподобия порождаемых гипотез. В работе [10] показано, что при некотором расширении QCA (дополнении используемых логико-алгебраических процедур поиска сходства адекватными процедурами вывода по аналогии и абдуктивного объяснения) оба варианта могут рассматриваться как реализация общей эвристической схемы «сходство – аналогия – абдукция».

Таким образом, формализованный качественный анализ социологических данных означает исследование психосоциальных явлений на более высоком уровне формализации и предполагает возможность первичной структуризации данных и знаний, формирование системы отношений для них и выбор адекватных формальных средств анализа. Очевидно, что реализация такого анализа возможна лишь в

компьютерных системах высокого уровня, содержащих средства извлечения знаний из баз фактов (БФ), автоматического порождения гипотез и объяснения имеющихся фактов на основании порожденных гипотез способных также осуществлять дедуктивный вывод из исходных и полученных знаний (базы знаний, БЗ). Именно и только системы, обладающие указанными возможностями, были охарактеризованы в [11] как интеллектуальные (ИС). Основанием для такой характеристики является фундаментальный тезис об основной задаче ИС как задаче конструктивной имитации (возможно, лишь до некоторой степени) познавательных способностей человека. Отметим, что многие средства анализа, традиционно рассматриваемые как *data mining*, оказываются не в состоянии выполнить эту сформулированную в [11] программу. Предлагаемое же представление о формализованном качественном анализе социологических данных как извлечении зависимостей из эмпирических данных средствами формализованных познавательных процедур, возможно, позволит приблизиться к ее решению (для ограниченного фрагмента некоторой конкретной предметной области).

1. Структура интеллектуальной системы

Архитектура системы, предназначенной для приближенного отображения познавательной деятельности исследователя – интеллектуальной системы (ИС), – представлена в [11] следующим образом: ИС = Решатель задач + Информационная среда + Интеллектуальный интерфейс. Общие характеристики каждой составляющей подробно рассмотрены и описаны в указанной работе, нам же предстоит сосредоточиться на специфических для рассматриваемой задачи особенностях.

«Сердцем» ИС, разумеется, является Решатель задач – именно здесь реализуются познавательные процедуры, однако без развитых инструментов накопления знаний (Информационная среда), а также восприятия и отображения (Интеллектуальный интерфейс) познавательный процесс не может быть полноценным. Наш подход ориентирован на интеллектуальные системы типа ДСМ, составная часть Реша-

³ Иногда к интеллектуальному анализу данных относят также методы *data mining* [4], но это представляется некоторым преувеличением возможностей таких подходов.

⁴ Метод назван по инициалам английского философа и логика Д.С. Милля, описавшего фигуры индуктивных рассуждений.

теля которых – Рассуждатель – реализует процедуры ДСМ-метода автоматического порождения гипотез, формализующие соответствующую эвристику анализа данных. Однако, независимо от Решателя, ИС эффективно решает задачи ФКАСД, если при сохранении описанной структуры превращается в автоматизированное рабочее место социолога, обеспечивающее проведение качественного исследования. Сюда входит хранение, структуризация и управление данными (препроцессинг) в едином информационном пространстве, применение различных стратегий анализа данных, реализуемых Решателем (ДСМ-метод или QCA), использование полученных результатов для последующих форм анализа (например, анализа рациональности мнений – об этом см. ниже). Отметим, что препроцессинг – подготовка данных и уточнение модели предметной области – может осуществляться как автоматически (с помощью специальных диагностических процедур, позволяющих выбрать наиболее адекватные стратегии Решателя), так и в интерактивном режиме.

В технологическом смысле можно рассматривать Информационную среду не просто как совокупность БФ и БЗ (как это сделано в [11]), но как активную оболочку ИС, обеспечивающую функционирование всех компонентов системы как единого целого. Здесь не только осуществляется надежное и безопасное хранение и управление данными, но и организуются и упорядочиваются процессы взаимодействия пользователей с Решателем в целях максимально эффективного использования возможностей предлагаемых технологий качественного анализа данных. Интеллектуальный интерфейс обеспечивает возможность внесения изменений в Информационную среду, предоставляет различные инструменты препроцессинга и предлагает средства интеграции с другими прикладными системами.

Последнее обстоятельство особенно важно в контексте рассматриваемой задачи анализа социологических данных, где огромную роль играют укоренившиеся традиции обработки результатов эмпирических исследований. Одним из наиболее широко распространенных инструментов такого анализа является пакет SPSS –

Statistical Package for the Social Sciences (для обучения работе с которым существуют специальные издания – см., например, [12]). Соответственно, практические социологи собирают и накапливают эмпирический материал, по большей части, в формате SPSS, и таковой материал должен быть доступен для ИС, реализующих ФКАСД. Далее, социологу могут понадобиться инструменты не только качественного, но и количественного анализа, и в этом случае использование средств SPSS оказывается более чем оправданным. Отметим, что в таком варианте ИС может характеризоваться как специфическая интегрированная [11]: наряду с Рассуждателем Решатель включает также *внешний* Вычислитель и, соответственно, Синтезатор, регулирующий взаимодействие Рассуждателя и Вычислителя.

Вариант системы, отвечающей целям реализации ФКАСД и соответствующий указанным требованиям, описан в [13, 14].

2. Класс решаемых задач

Познавательная деятельность, которая реализуется посредством интеллектуальной системы описанной выше архитектуры, характеризуется в [11] как интеллектуальный анализ данных (ИАД). Соответственно, формализованный качественный анализ социологических данных, осуществляемый средствами соответствующей ИС, есть вид ИАД и, как таковой, призван решать основные проблемы когнитивной социологии. Крупнейший американский социолог Т. Парсонс [15] так определяет эти проблемы: «Во-первых, Вебер определял социологию как научную дисциплину, которая, в первую очередь, должна попытаться понять действия индивидов, особенно в их социальных отношениях. ...Во-вторых, Вебер считал, что социология помимо субъективных мотивов должна развивать каузальные объяснения процесса действия, его направления и последствий». Добавив к этому, в соответствии с идеями М. Вебера [16], проблему исследования рационального поведения и отклонений от него, мы получим круг задач ФКАСД.

Итак, интеллектуальная система типа ДСМ для анализа социологических данных должна решать следующие основные задачи:

– исследование индивидуального поведения, порождение детерминант поведения и типологизация социума на их основе;

– анализ и прогнозирование мнений респондентов как варианта поведения;

– выяснение влияния ситуации на поведение индивидуума;

– анализ рациональности мнений (в т.ч. степени рациональности мнений данной социальной общности).

На успешное решение этих задач можно надеяться лишь при выборе методов анализа, адекватных исследуемой социальной реальности⁵. Для достижения этой цели требуется описание онтологических допущений относительно типов исследуемых предметных областей [11]. Вопреки традициям количественного анализа, социальные явления по большей части не представляются множествами случайных событий, которые только и могут изучаться статистическими средствами. Так, на детерминистский характер большинства социологических явлений указывает К. Поппер: «причинные законы в социальных науках ... являются качественными, а не количественными и математическими. Если социологические законы и определяют степень чего-либо, то используют при этом весьма неопределенные понятия и в лучшем случае дают очень грубую оценку» [18]. Соответственно, анализ такого рода событий должен осуществляться детерминистскими методами. Разумеется, о причинной обусловленности социальных явлений (таких как, к примеру, индивидуальное поведение) можно говорить, имея в виду некоторые уточнения. Детерминированность здесь – это, скорее, предрасположенность (в смысле К. Поппера) к совершению поведенческих актов (действий, установок, мнений). Более того, предрасположенность эта реализуется при отсутствии противодействующих влияний (как внутренних – личностных, так и внешних – ситуационных).

Существенной особенностью рассматриваемой области является открытость знаний: неслучайно одним из краеугольных камней подхода *обоснованной* теории (см. введение) является необходимость пополнения информа-

ции. Соответственно, для решения задач ФКАСД требуется применение формализованных *правдоподобных* рассуждений, позволяющих выдвигать гипотезы (которые соответствуют БФ) с конструктивно порождаемыми степенями правдоподобия. Заметим здесь, что метод формализованного качественного анализа QCA как таковой реализует логико-алгебраический вывод – построение дизъюнктивных нормальных форм на основании истинностной таблицы – и, стало быть, предполагает «замкнутость мира» (или, в терминах обоснованной теории, насыщение информации). Однако, как было показано в [10], метод может быть модифицирован для использования в «открытых мирах» в соответствии с эвристической схемой «сходство – аналогия – абдукция».

Другим принципиальным допущением является предположение о наличии как позитивных, так и негативных фактов – наличия или отсутствия исследуемого явления, вызванного позитивными (+)- и негативными (-)-причинами (наиболее существенными и устойчивыми влияниями), соответственно. Выполнение этого условия позволяет автоматически порождать фальсификаторы порожденных гипотез и может рассматриваться как основание для абдуктивного принятия индуктивных гипотез о причинах. Отметим, что в QCA это условие также является существенным, хотя и используется опосредованно – для описания значений *false* в истинностной таблице. «Миры» (социумы), которые удовлетворяют эти требованиям, будем обозначать $W^{(\pm)}$.

3. Структуризация данных.

Синтез познавательных процедур

Фундаментальным принципом качественного анализа данных является принцип «сходство фактов влечет наличие (отсутствие) изучаемого эффекта и его повторяемость». Конкретизируя его для реализации идеи причинности, можно сказать, что в рамках качественного анализа исследуется тип каузальности «структура – эффект» (а не «явление – явление»). Гипотетические причины представляются в виде сходств фактов, имеющих определенную структуру, т.е. «сходство» в этой схеме являет-

⁵ Об остроте проблемы такого соотношения писал еще П. Сорокин [17].

ся нестатистическим и конкретизируется посредством логико-алгебраического (QCA) и формально-индуктивного подхода (ДСМ-метод АПГ). Подчеркнем, что порождение причинно-следственных зависимостей на основе структурного сходства позволяет анализировать небольшие массивы данных (например, малые группы), что является несомненным преимуществом качественного анализа по сравнению с количественным при решении ряда задач.

Из описанного принципа структурализма вытекает потребность первичной (до решения собственно задач ФКАСД, описанных в предыдущем разделе) структуризации данных и знаний, т.е. предварительной алгебраической формализации сходства объектов и их свойств. Однако при существующей традиции обработки эмпирических социологических данных количественными методами и, соответственно, подготовки данных именно для такого анализа реализация этой потребности трудно достижима.

В предлагаемом варианте формализованного качественного анализа социологических данных с использованием ДСМ-рассуждений указанная трудность преодолевается выделением обозримого множества характеристик социальных субъектов (как индивидов, так и социальных общностей). Основой представления знаний о субъекте является так называемый «постулат поведения». Этот постулат адекватно отражает представления современных социологов: «для анализа поведения необходимо знать ... структуру характера, типичного для данной социальной системы» (Т. Парсонс, «О структуре социального действия», [15]). Сюда же можно отнести представления Д.С. Милля о возможности предсказания поведения индивида на основе знания о его характере и настроениях.

Пусть имеются три множества характеристик, входящих в описание субъекта поведения: признаки, представляющие социальный характер субъекта (SC); индивидуальные черты личности (IP); биографические данные (BD). Поведение B субъекта C определяется подмножеством характеристик $Det \subseteq C$ таким, что $Det = Det_1 \cup Det_2 \cup Det_3$, где $(Det_1 \subseteq (SC)) \& (Det_2 \subseteq (IP)) \& (Det_3 \subseteq (BD))$, причем хотя бы одно $Det_i \neq \emptyset$, $i = 1, 2, 3$. Таким образом, индивидуальные характеристики социаль-

ного субъекта являются информативным основанием для порождения детерминант социального поведения и, соответственно, материалом для построения возможных моделей социальной структуры с использованием установленных детерминант поведения.

Итак, формализуемость отношения сходства между позитивными (+)- и негативными (-)-фактами (далее иногда – (\pm)-фактами) мира (социума) $W^{(\pm)}$, наличие в БФ позитивных и негативных примеров изучаемого эффекта поведения и предполагаемая (\pm)-квасисимметрия причин (существенных влияний)⁶ проявления этих эффектов являются основанием для реализации основного принципа качественного анализа социологических данных. Указанные три фундаментальных допущения описывают класс задач, для решения которых создаются различные ИС, реализующие один тип формализованных эвристик: «сходство – аналогия – абдукция».

В ДСМ-методе автоматического порождения гипотез указанная схема формализованного качественного анализа социологических данных представляется эвристической схемой «индукция – аналогия – абдукция» и конкретизируется синтезом познавательных процедур: эмпирической индукции (основанной на установлении сходства фактов и представляющей формальные расширения и уточнения индуктивных методов английского философа и логика Д.С. Милля), каузальной аналогии (перенос гипотез о причинах на случаи с неопределенным исходом) и абдукции Ч.С. Пирса, посредством которой объясняется начальное состояние базы фактов и, в случае необходимости, осуществляется ее пополнение. Такой синтез, примененный к объединению БФ и БЗ, порождает новое знание – расширение БЗ. Как показано в [11], в результате работы соответствующих ИС из данных не только извлекаются «образцы» в смысле data mining, но и порождаются гипотезы о закономерностях, которые обнаруживаются в расширяемых (вследствие потребностей абдуктивного объяснения имеющихся фактов) последовательностях БФ.

⁶ Мы говорим о квазисимметрии (а не о полной симметрии) позитивных и негативных влияний, так как отсутствие эффекта может быть вызвано не только наличием негативного влияния, но и просто отсутствием позитивного.

Подчеркнем, что наличие в социуме причинно-следственных (\pm)-зависимостей типа «подмножество характеристик субъекта есть причина наличия (отсутствия) эффектов поведения (поведенческих актов, установок, мнений)» не просто делает качественный анализ предпочтительнее статистического (количественного). Это фундаментальная характеристика социума $W^{(\pm)}$, которая является основанием для реализации двух важнейших функций формализованного эмпирического исследования. Во-первых, наличие (\pm)-причин позволяет автоматически порождать фальсификаторы индуктивных гипотез, что увеличивает их степень правдоподобия (напомним, что Рассуждатель ДСМ-системы реализует *правдоподобные* рассуждения). Во-вторых, представление о наличии у каждого (\pm)-факта объясняющей его (\pm)-причины (называемое в ДСМ-методе аксиомами каузальной полноты АКП^(\pm)) является основанием для абдуктивного принятия порождаемых гипотез.

Общая схема ДСМ-рассуждения для формализованного качественного анализа социологических данных может быть представлена следующим образом.

В исходном состоянии БФ представлены предикаты $X \Rightarrow_1 Y$, которые интерпретируются как «субъект X обладает эффектом поведения Y », где X – структурированное описание субъекта (например, в соответствии с «постулатом поведения»), Y – переменная для представления действий, установок и мнений. Производные предикаты $V \Rightarrow_2 W$ и $W \Leftarrow_3 V$ означают, что «подмножество характеристик V есть причина эффекта поведения W » и «эффект поведения W есть следствие подмножества характеристик V », соответственно.

ДСМ-рассуждение состоит из последовательного и итерируемого применения индуктивных выводов (из предиката $X \Rightarrow_1 Y$ порождаются предикаты $V \Rightarrow_2 W$ или $W \Leftarrow_3 V$, т.е. в автоматическом режиме формируются фрагменты базы знаний интеллектуальной системы типа ДСМ) и выводов по аналогии (они используют гипотезы $V \Rightarrow_2 W$ и $W \Leftarrow_3 V$ о причинах изучаемых эффектов, порожденные индукцией, для расширения и уточнения представленного в начальном состоянии базы фактов отношения

\Rightarrow_1^*). Цикл «индукция – аналогия» повторяется до стабилизации множества гипотез $H = H_1 \cup H_2$, где H_1 – гипотезы о причинах изучаемых эффектов, полученные с использованием правил правдоподобного вывода 1-го рода – индукции, а H_2 – гипотезы, являющиеся предсказаниями и полученные с использованием правил правдоподобного вывода 2-го рода – аналогий. ДСМ-рассуждение завершается применением абдукции (формализованной посредством критерия достаточного основания принятия гипотез) – процедуры объяснения начального состояния БФ.

4. Эволюционная эпистемология

В работе [11] показано, каким образом указанный процесс проецируется на известную схему роста знания по К. Попперу [19]: $P1 \rightarrow TT \rightarrow EE \rightarrow P2$. Здесь $P1$ – рассматриваемая проблема, TT – пробная теория, EE – коррекция пробной теории и, наконец, $P2$ – новая проблема, возникающая после анализа скорректированной теории. Можно добавить несколько замечаний, которые позволяют прояснить некоторые существенные различия качественного и традиционного статистического анализа социологических данных. Статистический анализ часто используется для проверки заранее сформулированных гипотез, т.е. шаг $P1 \rightarrow TT$ фактически является априорным и эмпирические данные используются лишь на следующем шаге $TT \rightarrow EE$, когда теория проверяется и корректируется. Однако собственно переход от $P1$ к TT – извлечение скрытых закономерностей (т.е. знаний) – и является ключевой проблемой анализа неформализованных эмпирических данных. Недаром создатели «обоснованной теории» [3] – специально созданной (неформальной) методологии качественного анализа социологических данных – считают этот процесс краеугольным камнем своего подхода, подчеркивая, что он остается за рамками количественного анализа. Соответственно, ДСМ-метод как метод формализованного качественного анализа социологических данных полностью его обеспечивает.

Коррекция EE пробной теории TT в статистическом анализе сводится к непосредственной

проверке *выдвинутых* гипотез на массиве эмпирических данных. В ДСМ-методе процедура ТТ→ЕЕ существенно шире. На этом этапе исследования формируется абдуктивное объяснение *порожденных* (на основе БФ) гипотез и, при необходимости, коррекция и расширение БФ. Здесь же возможно изменение стратегии Рассуждателя. Расширение БФ и изменение стратегии для уточнения теории ТТ контролируется специально введенной характеристикой – степенью абдуктивной полноты [8, 11]: отношением числа (\pm)-фактов из БФ, объясненных порожденными (\pm)-гипотезами, к общему числу фактов.

Рассмотрим последовательность расширяемых баз фактов $B\Phi_1 \subset B\Phi_2 \subset \dots \subset B\Phi_m$. Пусть $B\Phi_i^+$ и $B\Phi_i^-$, соответственно, – множества (+)- и (-)- фактов из БФ, $i = 1, \dots, m$, $|B\Phi_i^\sigma|$ – число элементов $B\Phi_i^\sigma$, $\sigma \in \{+, -\}$. Пусть, далее, $\overline{B\Phi}_i^\sigma \subseteq B\Phi_i^\sigma$, где $\overline{B\Phi}_i^\sigma$ – подмножество $B\Phi_i^\sigma$ такое, что для каждого (σ)-факта из этого подмножества найдется объясняющая его (σ)-гипотеза (выше упоминалось, что в ДСМ-методе это называется выполнением аксиомы каузальной полноты АКП^(σ)). Тогда, если существует m для назначенного порога ρ^σ абдуктивного принятия гипотез такое, что $\rho_{i1}^\sigma \leq \dots \leq \rho_{im}^\sigma$

$\geq \rho^\sigma$, где $\rho_{i\sigma}^\sigma = \frac{|\overline{B\Phi}_i^\sigma|}{|B\Phi_i^\sigma|}$, $i = 1, \dots, m$, $\sigma \in \{+, -\}$, то

будем говорить, что имеет место абдуктивная сходимость ДСМ-метода АПГ.

Наконец, этап ЕЕ→Р2 в явном виде, как правило, вообще не присутствует в традиционном статистическом исследовании эмпирических социологических данных. В рамках же интеллектуального анализа данных скорректированная теория может приводить исследователя к необходимости решения новой проблемы Р2. Так, в рамках рассматриваемого ФКАСД проблема Р2 может формулироваться как решение задачи структуризации данных типа «социальная общность – эффекты поведения» и создание средств анализа и прогнозирования эффектов поведения для таких данных [20, 21]. Основанием для построения модели структуры изучаемого социума являются порожденные ДСМ-методом АПГ детерминанты поведения (действий, установок, мнений), полученные в

результате решения проблемы Р1 – представления и анализа знаний об индивидуальном поведении субъекта. Разумеется, дальнейший анализ социальных общностей необязательно ограничивается предложенной характеристикой таких – так, внутри общностей могут быть заданы отношения, сами общности могут также быть связаны различными отношениями и т.п.

5. Краткое изложение ДСМ-метода автоматического порождения гипотез

Рассмотрим теперь (со степенью подробности, достаточной для наших целей⁷) используемые нами формальные средства качественного анализа социологических данных. ДСМ-метод состоит из формального языка с дескриптивной и аргументативной функциями, правдоподобных ДСМ-рассуждений (синтеза трех познавательных процедур), квазиаксиоматических теорий КАТ для представления и систематизации открытого множества знаний о предметной области, дедуктивной имитации правдоподобных рассуждений в КАТ (гарантирующей корректность их результатов) и, наконец, интеллектуальных систем с Рассуждателем, реализующим ДСМ-рассуждения, и базой знаний БЗ, представленной соответствующей КАТ.

Массив начальных данных содержит высказывания типа «высказывание «субъект С обладает эффектом поведения Q» имеет истинностное значение $\langle v, n \rangle$ » ($J_{\langle v, n \rangle}(C \Rightarrow_1 Q)$ в ДСМ-языке). Здесь $v \in \{1, -1, 0, \tau\}$ – типы истинностных значений «фактическая истина», «фактическая ложь», «фактическое противоречие» и «неопределенность», соответственно, n – номер шага вычислений, отражающий степень правдоподобия истинностного значения. $J_{\langle v, n \rangle}\Phi = t$, если $v[\Phi] = \langle v, n \rangle$; $J_{\langle v, n \rangle}\Phi = f$, если $v[\Phi] \neq \langle v, n \rangle$, $v[\Phi]$ есть функция оценки, $\langle v, n \rangle$ представляет «внутренние» истинностные значения фактов и гипотез, t, f – «внешние» истинностные значения двузначной логики. $\langle v, 0 \rangle$ – истинностное значение факта, $\langle v, n \rangle$ – истинностное значение

⁷ Полное и систематическое изложение ДСМ-метода автоматического порождения гипотез можно найти в [9, 22].

гипотезы при $n > 0$. Для истинностных значений бесконечнозначных ДСМ-логик имеет место рекуррентное соотношение: $(\tau, n) = \{\langle 1, n+1 \rangle, \langle -1, n+1 \rangle, \langle 0, n+1 \rangle\} \cup (\tau, n+1)$.

В результате применения ДСМ-рассуждений порождаются гипотезы вида $J_{\langle v, n \rangle}(C' \Rightarrow_2 Q)$ или $J_{\langle v, n \rangle}(C' \Leftarrow_3 Q')$ (в зависимости от используемой стратегии анализа данных, о чем мы будем говорить ниже), $n > 0$. $J_{\langle v, n \rangle}(C' \Rightarrow_2 Q)$ означает, что «высказывание “подмножество характеристик субъекта C' есть причина эффекта поведения Q ” имеет истинностное значение $\langle v, n \rangle$ ». Соответственно, $J_{\langle v, n \rangle}(C' \Leftarrow_3 Q')$ означает, что «высказывание “эффект поведения Q' есть следствие характеристик субъекта C' ” имеет истинностное значение $\langle v, n \rangle$ ». Таким образом, предикаты $V \Rightarrow_2 W$ и $W \Leftarrow_3 V$ представляют порожденное из предиката $X \Rightarrow_1 Y$ отношение причинности. Таким образом, в этом процессе в автоматическом режиме формируются фрагменты базы знаний интеллектуальной системы типа ДСМ.

Высказывания вида $J_{\langle v, 0 \rangle}(C \Rightarrow_1 Q)$ суть факты, $J_{\langle v, n \rangle}(C \Rightarrow_j Q)$ ($j = 1, 2$) и $J_{\langle v, n \rangle}(C' \Leftarrow_3 Q')$, $n > 0$, – гипотезы.

Пусть даны конечные множества $U^{(1)} = \{d_1, \dots, d_r\}$ – множество характеристик социальных субъектов (описанных в соответствии с постулатом поведения) и $U^{(2)} = \{a_1, \dots, a_s\}$ – множество эффектов их поведения. Тогда структуризация знаний и фактов о социуме состоит в задании двух булевых алгебр: $B_1 = \langle 2^{U^{(1)}}, \emptyset, U^{(1)}, \text{---}, \cap, \cup \rangle$ – алгебры субъектов поведения и $B_2 = \langle 2^{U^{(2)}}, \emptyset, U^{(2)}, \text{---}, \cap, \cup \rangle$ – алгебры эффектов поведения (действий, установок, мнений) и предикатов – $\Rightarrow_1: 2^{U^{(1)}} \times 2^{U^{(2)}} \rightarrow V$; $\Rightarrow_2: 2^{U^{(1)}} \times 2^{U^{(2)}} \rightarrow V$; $\Leftarrow_3: 2^{U^{(2)}} \times 2^{U^{(1)}} \rightarrow V$, где $V = \{\langle v, n \rangle | v \in \{1, -1, 0, \tau\}, n \in N\}$, N – множество натуральных чисел. Определения переменных и констант сортов 1 и 2 – объектов $X \in 2^{U^{(1)}}$ и множеств свойств $Y \in 2^{U^{(2)}}$, соответственно, а также термов и формул ДСМ-языка содержатся в [22].

ДСМ-рассуждение формализуется средствами бесконечнозначной логики предикатов (1-го

порядка для конечных моделей и слабой логики предикатов 2-го порядка для бесконечных моделей [23]).

Первый шаг ДСМ-рассуждения – порождение гипотез о причинах вида $J_{\langle v, n \rangle}(C' \Rightarrow_2 Q)$, где $n > 0$, $v \in \{1, -1, 0\}$, или $J_{\langle \tau, n \rangle}(C' \Rightarrow_2 Q)$ на основании правил индуктивного вывода п.п.в.-1 (правил правдоподобного вывода 1-го рода). В этом варианте индукции – прямом методе сходства – реализуется условие «сходство субъектов поведения (предполагаемая причина) влечет наличие эффектов поведения».

Для формализации индукции определяются предикаты простого сходства $M_{a,n}^+(V, W)$ и $M_{a,n}^-(V, W)$ [22], где n – параметр, отображающий число применений правил правдоподобного вывода к БФ. $M_{a,n}^+(V, W)$ служит для порождения утверждений о возможных причинах наличия изучаемых свойств. Предикат включает подформулы, содержащие экзистенциальное условие (существование k (+)-фактов вида

$$J_{(1,n)}(X_i \Rightarrow_1 Y_i), i = 1, \dots, k, J_{(1,n)}\Psi = \bigvee_{j=1}^n J_{(1,j)}\Psi); \text{условие сходства (+)-фактов } (\bigcap_{i=1}^k X_i = V) \& (V \neq \emptyset);$$

условие причинно-следственной зависимости V и W и условие исчерпываемости сходных (+)-фактов $\forall X \forall Y ((J_{(1,n)}(X \Rightarrow_1 Y) \& (V \subset X)) \rightarrow \rightarrow ((W \subset Y) \& (\bigvee_{i=1}^k (X = X_i))))$; условие нижней границы числа сходных (+)-фактов $k \geq 2$ (она может быть увеличена). Аналогично определяется $M_{a,n}^-(V, W)$ для порождения кандидатов в гипотезы о причинах (–)-фактов.

Собственно гипотезы о (+)-причинах есть результат применения правил п.п.в.-1, имеющих следующий вид:

$$(I)_n^+ \frac{J_{\langle \tau, n \rangle}(V \Rightarrow_2 W), M_{a,n}^+(V, W) \& \neg M_{a,n}^-(V, W)}{J_{\langle 1, n+1 \rangle}(V \Rightarrow_2 W)},$$

где $J_{\langle \tau, n \rangle}(V \Rightarrow_2 W)$ – посылка (на n -м шаге неизвестно, является ли V причиной W), а $J_{\langle 1, n+1 \rangle}(V \Rightarrow_2 W)$ – заключение, которое является гипотезой о (+)-причине с истинностным значением $\langle 1, n+1 \rangle$ и со степенью правдоподобия, характеризующейся номером шага рассужде-

ний $n+1$. Аналогично формулируются $(I)_n^-$, $(I)_n^0$ и $(I)_n^\tau$ для заключений $J_{(-1,n+1)}(V \Rightarrow_2 W)$, $J_{(0,n+1)}(V \Rightarrow_2 W)$ и $J_{(\tau,n+1)}(V \Rightarrow_2 W)$ и посылок $\neg M_{a,n}^+(V,W) \& M_{a,n}^-(V,W)$, $M_{a,n}^+(V,W) \& M_{a,n}^-(V,W)$ и $\neg M_{a,n}^+(V,W) \& \neg M_{a,n}^-(V,W)$, соответственно.

Возможна и иная стратегия индуктивного анализа, определяющая *обратный* метод схождения, когда реализуется принцип «сходство эффектов поведения определяет сходство самих субъектов поведения». Для этого варианта с некоторыми модификациями определяются предикаты обратного схождения $\tilde{M}_{a,n}^\sigma(V, W)$ [20] ($\sigma \in \{+, -\}$), которые содержат подформулы, описывающие условия, подобные условиям $M_{a,n}^\sigma(V,W)$. Гипотезы, порождаемые соответствующими п.п.в.-1 $(\tilde{I})_n^\sigma$, имеют вид $J_{(v,n+1)}(W \Leftarrow_3 V)$ или $J_{(\tau,n+1)}(W \Leftarrow_3 V)$ (« W – следствие причин V »), $v \in \{1, -1, 0\}$, $\sigma \in \{+, -, 0, \tau\}$.

Таким образом, отношение, представленное в БФ предикатом $X \Rightarrow_1 Y$, посредством $(I)_n^\sigma$ и $(\tilde{I})_n^\sigma$ ($\sigma \in \{+, -, 0, \tau\}$) порождает отношения \Rightarrow_2^* и \Leftarrow_3^* , соответственно. Эти отношения представлены множеством гипотез H_1 , являющимся фрагментом БЗ.

Следующий шаг ДСМ-рассуждения – вывод по аналогии – выполняется посредством правил правдоподобного вывода 2-го рода (п.п.в.-2). Для этих правил формулируются предикаты $\Pi_n^\sigma(V, W)$ ($\sigma \in \{+, -, 0, \tau\}$), позволяющие предсказывать наличие (отсутствие) эффекта поведения W у субъекта V .

$\Pi_n^+(V, W)$ содержит подформулы, выражающие наличие в БЗ (+)-причин $J_{(1,n)}(X_i \Rightarrow_2 Y_i)$, $i = 1, \dots, k$, таких, что $X_i \subset V$, $\bigcup_{i=1}^k Y_i = W$; при этом не существует Z и U ($U \subseteq W$) таких, что $J_{(-1,n)}(Z \Rightarrow_2 U)$ (Z – (-)-причина U) и $Z \subset V$. Тогда применимо правило п.п.в.-2

$$(II)_n^+ \frac{J_{(\tau,n)}(V \Rightarrow_1 W), \Pi_n^+(V, W)}{J_{(1,n+1)}(V \Rightarrow_1 W)}$$

Аналогично определяются $\Pi_n^-(V, W)$ и $\Pi_n^0(V, W)$. $\Pi_n^\tau(V, W) = \neg(\Pi_n^+(V, W) \vee \Pi_n^-(V, W) \vee \Pi_n^0(V, W))$.

Для обратного ДСМ-метода АППГ предикаты $\tilde{\Pi}_n^\sigma(V, W)$ и правила $(\tilde{II})_n^\sigma$ ($\sigma \in \{+, -, 0, \tau\}$) с

использованием $J_{(v,n)}(X_i \Leftarrow_3 Y_i)$ формулируются аналогично. Гипотезы, порожденные посредством $(II)_n^\sigma$ и $(\tilde{II})_n^\sigma$, образуют множество H_2 , составляющее часть БЗ.

Из приведенного описания видно, что посредством отношений \Rightarrow_2^* и \Leftarrow_3^* (гипотез из H_1) порождается отношение \Rightarrow_1^* (гипотезы из H_2), расширяющее и уточняющее представленное в начальном состоянии БФ отношение, таким образом, гипотезы из H_1 являются аргументами для порождения гипотез из H_2 : аргументация формализуется посредством $\Pi_n^\sigma(V, W)$ или $\tilde{\Pi}_n^\sigma(V, W)$. Порожденные же гипотезы из H_2 аналогичны (в определенном смысле) «родителям» аргументов из H_1 , т.е. $J_{(v,n)}(X_i \Rightarrow_2 Y_i)$ (или $J_{(v,n)}(Y_i \Leftarrow_3 X_i)$). Напомним, что п.п.в.-1 и п.п.в.-2 применяются последовательно до стабилизации, когда новые гипотезы о причинах не порождаются: $H_{1n} = H_{1(n+2)}$, где $n, n+2$ – номера состояний БЗ ($n > 0$) и $H_{1n} = H_1$.

Схема абдуктивного объяснения – принятия гипотез – может быть представлена следующим образом.

БФ – множество фактов, представляющих $X \Rightarrow_1 Y$,

$H = H_1 \cup H_2$ – множество гипотез, порожденных индукцией и аналогией,

$E(H_1, \text{БФ})$ – реализация отношения объяснения БФ посредством H_1 .

$$\forall h_j ((h_j \in H) \rightarrow h_j \text{ правдоподобна}),$$

где $j = 1, \dots, 6$, h_j – гипотезы, $h_1 = J_{(v,n)}(C \Rightarrow_1 Q)$, $h_2 = J_{(v,n)}(C' \Rightarrow_2 Q)$, $h_3 = J_{(v,n)}(Q \Leftarrow_3 C')$, $h_4 = J_{(\tau,n)}(C \Rightarrow_1 Q)$, $h_5 = J_{(\tau,n)}(C' \Rightarrow_2 Q)$, $h_6 = J_{(\tau,n)}(Q \Leftarrow_3 C')$, а $v \in \{1, -1, 0, \tau\}$, $n > 0$. n – число применений правил правдоподобного вывода (п.п.в.-1 – индукции и п.п.в.-2 – аналогии), выражающее степень правдоподобия гипотезы h_j : чем больше n , тем меньше степень правдоподобия h_j .

Реализация отношения объяснения БФ порожденными гипотезами о причинах H_1 формализуется посредством упоминавшихся выше аксиом каузальной полноты (АКП^(±)) предметной области (социума) $W^{(\pm)}$: «всякий эффект социального поведения Y субъекта X имеет причины V_1, \dots, V_k , его вызывающие».

$$\text{АКП}^{(+)}: \forall X \forall Y \exists V_1 \dots \exists V_k \exists W_1 \dots \exists W_k \exists n_1 \dots \exists n_k$$

$$((J_{\langle 1, 0 \rangle}(X \Rightarrow_1 Y) \rightarrow ((\bigwedge_{i=1}^k (J_{\langle 1, n_i \rangle}(V_i \Rightarrow_2 W_i) \&$$

$$(V_i \subset X) \& (V_i \neq \emptyset) \& (W_i \neq \emptyset)) \& (\bigcup_{i=1}^k W_i = Y))).$$

АКП⁽⁻⁾ формулируется аналогично, поскольку, как уже говорилось, онтологической особенностью рассматриваемой социальной реальности является представление о (±)-квзисимметрии причин. Соответственно, формулируются А $\tilde{\text{КП}}^{(\pm)}$ для обратного метода [20].

Структурированное представление данных о субъекте и его поведении, конструктивное порождение гипотез о причинах тех или иных эффектов социального поведения и его прогнозирование, формализация объяснения эмпирических фактов с использованием аксиом каузальной полноты означает, что предложенными средствами осуществляется не только анализ данных, но и формирование новых знаний о предметной области и их систематизация посредством порождения новых отношений из исходных. Таким образом, мы действительно имеем вариант формализованного качественного анализа социологических данных, как это было определено во введении.

6. Семантика анализа мнений

Как уже говорилось, возможны две стратегии, различающиеся направленностью индуктивного анализа данных: «от причины – к следствию» (прямой) и «от следствия – к причине» (обратный). Выбор стратегии зависит от представления данных о субъекте и его поведении. При изучении собственно поведения (действия) или установок субъекта информативность представленных данных о субъекте, как правило, превосходит информативность данных о его поведении. В этом случае используется *прямой* ДСМ-метод.

При решении задачи анализа мнений, напротив, информативность характеристики мнения превосходит информативность знаний о субъекте, высказывающем мнение. Отсюда возникает потребность в формализации *обратного* рассуждения, анализирующего сходство мнений субъектов и на основании этого анализа выяв-

ляющего сходство самих субъектов, имеющих общие мнения. Ключевым вопросом при этом является возможность структурированного представления мнений (подобно тому, как было структурировано описание индивидуумов на основании постулата поведения).

Первым шагом к такой структуризации является для нас идея немецкого социолога Н. Лумана о том, что опросы общественного мнения должны быть представлены как ответы на вопросы по соответствующей теме [24]. На основе такого представления в [20] предложен вариант семантики ДСМ-метода для анализа и прогнозирования мнений⁸. Пусть задана некоторая тема опроса T*, характеризующаяся утверждениями из множества P = {p₁, ..., p_n}, которое будем называть каркасом темы T*, а элементы p₁, ..., p_n – корнями вопросов (параметрами опроса). В результате опроса устанавливается отношение респондентов к элементам каркаса и к теме в целом. Задана функция оценки v[p_j] с областью значений {1, -1, 0, τ}. Каждому элементу p_j (j = 1, ..., n) каркаса P соответствует вопрос ? p_j – «Какова оценка v корня вопроса p_j ?», v ∈ {1, -1, 0, τ}, ответом же является высказывание J_vp_j. J_vp_j = t, если v[p_j] = v; J_vp_j = f, если v[p_j] ≠ v.

Тогда ответом j-го респондента по теме T* будем называть максимальную конъюнкцию φ_j = J_{v₁^(j)} p₁ & ... & J_{v_n^(j)} p_n (вид мнения), где φ_j – метасимвол, “=” – предикат графического равенства формул, v_i^(j) ∈ {±1, 0, τ}, i = 1, ..., n; j = 1, ..., 4ⁿ. Такой ответ представляет собой понимание j-м респондентом темы T*. Множество членов этой конъюнкции обозначим [φ_j] = {J_{v₁^(j)} p₁, ..., J_{v_n^(j)} p_n} и будем называть составом мнения.

Множество всех возможных ответов по теме T* с каркасом P обозначим K.

K = {φ_j | φ_j = {J_{v₁^(j)} p₁ & ... & J_{v_n^(j)} p_n}, j = 1, ..., 4ⁿ, |K| = 4ⁿ. Заметим, что число респондентов может превышать 4ⁿ, поскольку различные респонденты могут иметь одинаковые ответы, при этом число различных ответов может быть меньше 4ⁿ.

⁸ В [25] предложено ее расширение для m-значного опроса (m ≥ 3).

Предлагаемая семантика позволяет формализовать анализ мнений средствами ДСМ-метода АПГ в соответствии со следующей стратегией. Формулируется тема мнения, пригодная для построения модели социальной структуры. Затем задается система вопросов, раскрывающих содержание темы – каркас темы. Оценка эмпирического отношения $C \Rightarrow_1 Q$ («субъект – мнение») есть оценка отношения к теме в целом, Q – состав мнения субъекта C (множество $\{J_{v_1} p_1, \dots, J_{v_n} p_n\}$ образующих (атомов) мнения с оценками, $Q = [\varphi]$).

Пусть $U^{(2)} = \{\psi \mid (\psi \neq J_{v_i} p_i) \& (v_i \in \{1, -1, 0, \tau\}), i = 1, \dots, n\}$, $|U^{(2)}| = 4n$. Задача изучения мнений сводится к изучению высказываний из БФ $J_{\bar{\mu}_j}(C_j \Rightarrow_1 [\varphi_j])$ – «субъект C_j имеет мнение φ_j ». В результате применения правил индуктивного вывода (п.п.в.-1) порождаются гипотезы вида $J_{\bar{\mu}_j}([\psi_j] \Leftarrow C'_j)$ – «мнение ψ_j есть следствие характеристик субъекта C'_j ». Порожденные детерминанты мнений в дальнейшем используются для прогнозирования мнений с помощью правил вывода по аналогии (п.п.в.-2), а также служат основанием для построения модели структуры изучаемого социума (и, соответственно, служат формулировке проблемы P2, о чем говорилось выше).

Здесь $C_j, C'_j, [\varphi_j], [\psi_j]$ – константы, $C_j, C'_j \in 2^{U^{(1)}}$, $[\varphi_j], [\psi_j] \in 2^{U^{(2)}}$, $|2^{U^{(2)}}| = 2^{|U^{(2)}|} = 2^{4n}$, $\bar{\mu}_j = \langle \mu, m \rangle$ – истинностное значение, полученное применением ДСМ-метода АПГ, где $\mu_j \in \{\pm 1, 0, \tau\}$, а m – число применений ДСМ-правил правдоподобного вывода.

Специфической особенностью анализа мнений является возможность конструктивного порождения насыщенной выборки – одной из основных идей обоснованной теории [4] – стратегии качественного анализа социологических данных. Пусть при опросе респондентов получено множество ответов $K' \subseteq K$, которое не изменяется при расширении множества опрашиваемых. В этом случае мы имеем дело со стабилизированным множеством ответов, причем стабилизация эта достигается экспериментальным путем. Рассмотрим множество респондентов $R = \{b_1, \dots, b_r\}$, соответствующее

множеству стабилизированных ответов K' , $R = \{X \mid \exists \varphi J_{(1,0)}(X \Rightarrow_1 [\varphi])\}$, $[\varphi] = \{J_{v_1} p_1, \dots, J_{v_n} p_n\}$, $\varphi \in K'$. Тогда множество респондентов R можно считать качественным аналогом репрезентативной выборки статистического анализа.

7. Ситуация как детерминирующий фактор поведения

Одной из перечисленных задач формализованного качественного анализа социологических данных является задача выяснения влияния ситуации на индивидуальное поведение. Согласно К.Р. Попперу ([19], Логика социальных наук, с. 298 – 314), введение знаний о ситуации сообщает социологическим рассуждениям объективный характер и устраняет психологизм как средство анализа социального поведения. Необходимость учета ситуации подтверждается современными исследованиями в области социальной психологии, а также результатами анализа социального поведения собственно средствами ДСМ-метода автоматического порождения гипотез [26].

Эти соображения послужили основой перехода от изучения отношения «субъект \Rightarrow поведение» к отношению « \langle субъект, ситуация $\rangle \Rightarrow$ поведение». Представленный в БФ бинарный предикат $X \Rightarrow_1 Y$ – «субъект X демонстрирует эффекты поведения Y » – заменяется тернарным предикатом $P(X, Y, S)$ – «субъект X демонстрирует эффекты поведения Y в ситуации S ». Предикат причинности, предусматривающий анализ сходства не только субъектов, но и ситуаций, по-прежнему остается бинарным, но причина является двухпараметрической, включающей ряд характеристик индивидуума (подобъект) и фрагмент ситуации (или всю ситуацию целиком). Традиционный предикат $V \Rightarrow_2 W$ («подмножество характеристик субъекта V есть причина эффектов поведения W ») заменяется предикатом $R_i(\langle V, S' \rangle, W)$ – «подмножество характеристик субъекта V и характеристики ситуации S' есть причина эффектов поведения W ». R_i ($i=1, 2, 3, 4$) характеризует структуру мира – насколько в ней существенны сам объект и ситуация проявления свойств. Пусть $R = R_1 \vee R_2 \vee R_3$. Тогда

$R_1 \leftrightarrow R \& (V \neq \emptyset \& S = \emptyset)$ (ситуация не влияет на поведение, как это и предусмотрено в простом варианте ДСМ-метода); $R_2 \leftrightarrow R \& (V \neq \emptyset \& S \neq \emptyset)$ (для анализа поведения существенны как личность индивидуума, так и ситуация поведения); $R_3 \leftrightarrow R \& (V = \emptyset \& S \neq \emptyset)$ (логика социальных наук К.Р. Поппера); $R_4 = R$ (реализуются все варианты причинности).

В соответствии с расширенным представлением о характере причинно-следственных зависимостей для ситуаций вводится новый сорт переменных $U^{(3)} = \{s_1, \dots, s_q\}$ – множество параметров, характеризующих ситуацию⁹. Соответственно, задается алгебра ситуаций (контекстов, внешних обстоятельств) $B_2 = \langle 2^{U^{(3)}}, \emptyset, U^{(3)}, \text{—}, \cap, \cup \rangle$.

Для ситуационного варианта ДСМ-метода определяются предикаты сходства ${}^3M_{a,n}^\sigma(V, W, S_0)$, где $i=1, 2, 3$ характеризует тип причинности, левый верхний индекс 3 указывает на тернарность сходства, $\sigma \in \{+, -\}$, n отображает число применений правил правдоподобного вывода к БФ. Предикат ${}^3M_{a,n}^+(V, W, S_0)$ включает подформулы, аналогичные подформулам предиката $M_{a,n}^+(V, W)$: экзистенциальное условие (существование k (+)-фактов вида $J_{(1,n)}P(X_j, Y_j, S_j), j = 1, \dots, k, k \geq 2$), условие сходства объектов и ситуаций, условие эмпирической зависимости и условие исчерпываемости сходных (+)-фактов $\forall X \forall Y \forall S ((J_{(1,n)}P(X, Y, S) \& (V \subset X) \& (S_0 \subseteq S)) \rightarrow (W \subseteq Y \& W \neq \emptyset \& (\bigvee_{j=1}^k (X = X_j))))$.

Симметрично формулируется предикат ${}^3M_{a,n}^-(V, W, S_0)$.

В результате применения соответствующих п.п.в.-1 $(I)^\sigma_i (\sigma \in \{+, -, 0\})$ или $(I)^\tau_i$ порождаются гипотезы о причинах $J_{(\epsilon, n+1)}R_i(\langle V, S \rangle, W)$ ($\epsilon \in \{+, -, 0\}$, соответственно), или $J_{(\tau, n+1)}R_i(\langle V, S \rangle, W)$ ($i=1, 2, 3$). Полученные гипотезы с помощью п.п.в.-2 переносятся на случаи недоопределенности в исходной БФ, т.е. с помощью соответствующих правил $(II)^\sigma_i (\sigma \in \{+, -, 0\})$ или $(II)^\tau_i$ (для кото-

рых определяются предикаты ${}_iK_n^\sigma(\langle V, S \rangle, W)$, $\sigma \in \{+, -, 0, \tau\}$ порождаются гипотезы $J_{(\epsilon, n+1)}P(V, W, S)$ ($\epsilon \in \{+, -, 0\}$, соответственно), или $J_{(\tau, n+1)}P(V, W, S)$ ($i=1, 2, 3$). Для абдуктивного принятия гипотез формулируются соответствующие аксиомы каузальной полноты ${}_iAKP^{(\sigma)}$, $\sigma \in \{+, -\}, i = 1, 2, 3$.

Для обнаружения влияния ситуации на формирование мнения вместо описанных *прямых* предикатов определяются *обратные* предикаты двух видов. Если исследователя интересует, какие особенности V субъекта определяют данное мнение W в ситуации S_0 , в результате применения п.п.в.-1, использующих решающие предикаты ${}_iM_{a,n}^\sigma(V, W, S_0)$, порождаются ги-

потезы вида $J_{(v, n)}\check{R}^p(V, \langle W, S_0 \rangle)$, $v \in \{1, -1, 0, \tau\}$, – «подмножество характеристик V субъекта определяет его мнение W в ситуации S_0 ». Если исследователя интересует, какое подмножество характеристик V субъекта и фрагмент ситуации S_0 есть причина наличия (отсутствия) мнения W – это задача собственно социологии. В этом случае использование решающих предикатов ${}_iM_{a,n}^\sigma(V, W, S_0)$ посредством п.п.в.-1 порождает гипотезы вида $J_{(v, n)}\check{R}^s(\langle V, S_0 \rangle, W)$ – «подмножество характеристик V субъекта и фрагмент ситуации S_0 есть причина наличия (отсутствия) мнения W »¹⁰.

Введение представления данных о ситуациях является существенным усилением предложенных логических средств формализованного качественного анализа социологических данных.

8. Анализ рациональности мнений

Анализ рационального поведения – классическое направление теоретической и прикладной социологии [28], охарактеризованное еще М. Вебером [16] как одно из центральных. Традиционное понимание рациональности опирается на представление о действии как акте, направленном на некоторую цель. В этом случае рациональность действия определяется как ус-

⁹ Разумеется, булево представление не является единственно возможным для описания ситуаций.

¹⁰ Описание соответствующих предикатов, правил правдоподобного вывода и аксиом каузальной полноты можно найти в [27].

танавливаемое с помощью рассудка соотношение между интенцией и имеющимися у субъекта знаниями, объективными возможностями и избранными средствами. Такой подход едва ли приложим, когда в центре исследовательского интереса оказывается проблема рациональности высказываемых респондентом мнений.

Описанная семантика ДСМ-метода АПГ для анализа мнений включает возможность описания и анализа такой рациональности, что является одной из сформулированных выше задач ФКАСД. При этом рациональность рассматривается вне зависимости от целеполагания субъекта (если таковое и имеется) как аргументированное принятие решений (высказываемых мнений), не сводимое к дедуктивному рассуждению [29].

Предложенное представление опроса по теме T с описанием каркаса темы $P = \{p_1, \dots, p_n\}$ есть приближение к аргументированному и, соответственно, рациональному восприятию темы. Возможность формирования рационального мнения может быть усилена, если предложить респонденту множество аргументов A относительно принятия или неприятия утверждений p_1, \dots, p_n [27]. Определим функции выбора $g^+(p_i): p_i \rightarrow 2^A$ и $g^-(p_i): p_i \rightarrow 2^A$ ($i = 1, \dots, n$), сопоставляющие каждому утверждению из P , соответственно, множество аргументов и контраргументов из A . Для изучения рационального мнения мы можем включить в описание j -го субъекта наряду с его дифференциальными признаками C_j также его аргументационную функцию, т.е. рассматривать предикат $\langle C_j, \langle g_j^+, g_j^- \rangle \rangle \Rightarrow_1 [\varphi_j]$, $j = 1, \dots, l$, где l – число рассматриваемых примеров. В идеальном случае, когда система аргументов A полна и каркас темы полностью раскрывает ее содержание, мнение $[\varphi_j]$ субъекта рационально, если оно полностью определяется его аргументационными функциями $g_j^+, g_j^-, g_j^\sigma = \{g_j^\sigma(p_1), \dots, g_j^\sigma(p_n)\}$, $\sigma \in \{+, -\}$, причем $g_j^+ \cap g_j^- = \emptyset$.

Однако, по М. Веберу, абсолютно рациональное поведение (мнение) реализуется лишь как поведение идеального типа. В реальности индивидуумы демонстрируют те или иные отклонения от идеального типа поведения. Предложенное представление мнений позволяет определить некоторые численные характеристики рациональности результатов опроса.

Одной из важнейших характеристик рациональности мнения является его непротиворечивость. Для ее распознавания в [27] предложен метод аналитических таблиц для логик JA_4 и JA_5 (4-х и 5-значных логик с аргументационной семантикой, соответственно, [27]). Рассмотрим непротиворечивое множество $\Sigma = \{\psi_1, \dots, \psi_s\}$ формул ψ_1, \dots, ψ_s логики JA_4 (JA_5), выражающих логические зависимости между элементами каркаса P (множество «постулатов значения», соответствующих теме T).

Пусть $consis(\Sigma \cup \{\varphi\})$ – метапредикат непротиворечивости множества формул $(\Sigma \cup \{\varphi\})$. Тогда множество всех φ (ответов респондентов) – максимальных конъюнкций формул $J_{v_i} p_i$ – противоречащих Σ , обозначим $\Delta = \{\varphi | \neg consis(\Sigma \cup \{\varphi\}) \& (\varphi \in K)\}$. Противоречивость мнения φ устанавливается простой процедурой: если аналитическая таблица \mathfrak{S} для множества $\Sigma \cup \{\varphi\}$ замкнута, то $\Sigma \cup \{\varphi\}$ – противоречиво и $\varphi \in \Delta$ (Δ – множество «запрещенных» максимальных конъюнкций для множества Σ).

Тогда степень непротиворечивости результатов опроса задается функцией $\delta(K', \Delta) = 1 - |K' \cap \Delta| / |K'|$. $\delta(K', \Delta) = 1$, если $K' \cap \Delta = \emptyset$ (опрос непротиворечивый); $\delta(K', \Delta) = 0$, если $K' \subseteq \Delta$ (опрос противоречивый); x , если $K' \cap \Delta \neq \emptyset$ и $\neg(K' \subseteq \Delta)$, $0 < x < 1$ (промежуточный опрос).

Степень непротиворечивости есть комплексная характеристика индивидуальных способностей респондентов различать заложенные исследователем связи между вопросами, описывающими заданную тему, и, разумеется, одинаковым образом вычисляется для различных по тематике опросов.

Рациональность мнения может задаваться и другими критериями: близостью к «идеальному мнению» и степенью согласованности мнений респондентов рассматриваемой социальной общности [30]. Первый характеризует степень сходства мнения респондента с некоторым «идеальным мнением», представленным «единственно правильным» ответом по теме T . Наиболее естественно этот критерий определяется для электоральной общности – в этом случае «идеальное мнение» сторонника некоей партии представлено принятием программных положений партии. Второй представляет близость

взглядов респондентов друг другу (их сходство) относительно некоторой темы и может касаться как электорального выбора, так и любых других тем, объединяющих респондентов в некоторую общность.

Особенностью предложенного описания рациональности мнений является возможность неоднократного ее вычисления, что расширяет содержательные возможности ФКАСД. Во-первых, характеристики рациональности вычисляются как для исходной БФ, так и для последовательно расширяющихся (ввиду потребностей индуктивного объяснения) БФ – тогда изменяется множество R и, соответственно, K' . Во-вторых, они вычисляются для результатов собственно ФКАСД (так называемых «предсказательных опросов») – когда формируются фрагменты БЗ, соответствующие доопределению фактов вида $J_{(т, о)}(C \Rightarrow_1[\Phi])$. Отметим, что предсказательный опрос характеризуется и другими объективными величинами – числом шагов применения эвристической схемы n (до стабилизации множества порождаемых гипотез), а также порогом абдуктивной сходимости ρ^σ , $\sigma \in \{+, -\}$.

9. Краткий обзор эмпирических результатов

Приведем примеры решения перечисленных задач формализованного качественного анализа социологических данных средствами ДСМ-метода автоматического порождения гипотез.

Задача первого типа – исследование индивидуального поведения, порождение детерминант поведения и типологизация социума на их основе – решалась на примере изучения солидарного поведения рабочих на конфликтных предприятиях (в частности, участии или неучастии в забастовке). Поскольку информативность представления субъекта в этом случае превосходит информативность описания его действий, в качестве стратегии анализа был выбран прямой ДСМ-метод АПГ.

Эмпирической базой эксперимента были данные, полученные сотрудниками лаборатории проф. В.А. Ядова Института социологии РАН при изучении рабочих двух предприятий: завода "Арсенал" в Санкт-Петербурге (157 респонден-

тов) и Завода тракторных гидроагрегатов в г. Ельце (132 респондента). Эти предприятия существенно отличались друг от друга по социокультурным параметрам, по состоянию трудовых отношений и стадии развития трудового конфликта (о наличии конфликтов на предприятиях было известно заранее) – фактически, по ситуационным параметрам. Впоследствии это сказалось на порождении различных детерминант различных форм коллективного поведения для этих предприятий, привело, соответственно, к различной типологизации рабочих коллективов и послужило практическим стимулом для создания ситуационного варианта ДСМ-метода.

Первичные (исходные) данные, непосредственно полученные как ответы на вопросы интервью и тестов, агрегировались и представлялись таким образом, что определенная комбинация ответов давала комплексную характеристику. Эти комплексные признаки конструировались исследователями на основании знаний о связях между переменными (полученных из опыта, методами факторного анализа и т.п.). Признаки были сформированы в соответствии с постулатом поведения:

- социальный характер – переменные, характеризующие обобщенные установки ценностного характера, нормативные установки, характеризующие предпочитаемую стратегию поведения в социально-трудовом конфликте;
- данные об индивидуально-психологических характеристиках испытуемых, полученные с помощью тестов, которые позволили выделить три психологических типа: лидеры, миссионеры, зависимые;
- биографические данные – ресурсные характеристики, обусловленные объективным статусом (материальное положение, образование, наличие приработков), характеристики идентичности (наемные работники или коллектив предприятия).

В качестве детерминируемых признаков рассматривались готовности к тому или иному виду поведения в проективной ситуации забастовки: *инициативный* (“войду в забастовочный комитет”), *активный* (“буду участвовать в забастовке”), *подражательский* (“поступлю как большинство”), *уход* (“не буду участвовать в забастовке”).

В результате анализа имеющихся данных средствами ДСМ-метода автоматического порождения гипотез были выявлены детерминанты различных видов поведения и на основе этих детерминант социологами определены различные поведенческие типы, причем, как уже говорилось, они оказались различными для изучаемых предприятий. Так, например, в Санкт-Петербурге был описан тип «Крестьянин, потерявший опору» (индивидуализм, непризнание закона в сочетании с желанием «договариваться», средний уровень материального положения). В Ельце – «Советский рабочий» (зависимость от предприятия, ощущение себя совладельцем, лояльность к директору, низкий уровень материального положения) [26].

Совместная работа оказалась плодотворной как для содержательного социологического исследования, так и для развития ДСМ-метода (идея аргументационного оценивания, увеличение числа типов истинностных значений – для промежуточных между +1 и –1 оценок, ситуационное расширение). Социологами были высоко оценены возможности комбинации социального и индивидуального, повышение уровня формализации данных и знаний, формирование эмпирических социологических понятий (типологизация), которые оказались основанием для уточнения социологической модели и порождения социологической концепции. Таким образом, ДСМ-метод оказался эффективным инструментом формализованного качественного анализа данных.

Эмпирическое исследование второго типа задач – анализ и прогнозирование мнений респондентов, а также анализ рациональности мнений – было осуществлено на массиве данных об электоральных предпочтениях студентов старших курсов РГГУ на выборах в Государственную Думу 2003 и 2007 гг. В работе принимали участие сотрудники, студенты и аспиранты Социологического факультета РГГУ, а также студенты и аспиранты Отделения интеллектуальных систем в гуманитарной сфере Института лингвистики РГГУ.

Для формализации качественного анализа данных о мнениях был выбран обратный ДСМ-метод. Решались следующие задачи.

1. порождение детерминант электорального поведения, представленного парой (мнение,

выбор действия), где «мнение» есть выбор программных установок (без указания в исходных данных их принадлежности конкретной партии), а «действие» – выбор одной из шести заданных партий или другой, или выбор «против всех», или отказ от участия в выборах. Иными словами, изучались расширенные высказывания $J_{\bar{d}_j}(C_j \Rightarrow_1 \langle [\Phi_j], T_i \rangle)$, где T_i – одна из выбираемых партий.

2. предсказание электорального выбора 27-ми (из 231-ти) опрошенных студентов (у которых был зафиксирован только выбор установок, но не электоральный выбор) посредством порожденных детерминант с последующей валидацией предсказаний посредством повторного опроса после голосования;

3. выяснение некоммутативности двух технологий опросов: а) сначала – мнение, затем – электоральный выбор, б) сначала – электоральный выбор, затем – мнение. С этой целью массив опрашиваемых разбивается на две равновеликие части;

4. анализ рациональности выбора программных установок и электорального действия (выбора политической партии в том числе).

Каркасом тем опроса являлись программные установки различных партий (по соответствующим проблемам – «земля», «приватизация», «СМИ», «внешняя политика», «экономика», «свобода личности», «социальная политика» и т.п.). Например (в исследовании 2003 г.), р₅₇: «Земля всецело должна быть в государственной собственности» (КПРФ); р₆₁: «Социальная рыночная экономика наиболее эффективна в условиях России» (Яблоко), р₆₈: «Необходимо узаконить итоги приватизации» (СПС), р₈₅: «Основная угроза России исходит от США» (ЛДПР).

В результате применения обратного ДСМ-метода были порождены, например, следующие зависимости (эксперимент 2007 г.):

$J_{\langle 1, 2 \rangle}(C \Rightarrow_1 \langle [\Phi], T_1 \rangle)$, где $C = \{Ж, \text{низкий уровень авторитарного подчинения, материальная помощь родителей, совмещение работы с учебой, бесплатное отделение, не замужем, средний уровень политической активности}\}$, $[\Phi] = \{J_{1p_1}, J_{1p_2}, J_{1p_3}, J_{-1p_4}, J_{0p_5}\}$, T_1 – Единая Россия. Здесь р₁: «государственная собственность должна преобладать над частной», р₂: «естественные монополии должны принадле-

жать государству», р₃: «армия должна быть не только контрактной, но и по призыву», р₄: «в формировании пенсии должно участвовать не только государство, но и работодатель», р₅: «внешняя политика России должна быть ориентирована на Запад» (здесь номера элементов каркаса не совпадают с номерами вопросов в анкете).

$J_{(1, 2)}(C \Rightarrow_1 \langle [\Phi], T_2 \rangle)$, где $C = \{4 \text{ курс, низкий уровень авторитарного подчинения, материальная помощь родителей, не женат/не замужем, средний уровень политической активности}\}$, $[\Phi] = \{J_{1p1}, J_{1p2}, J_{1p3}, J_{1p6}, J_{1p7}, J_{0p8}\}$, T_2 – СПС или Яблоко. Здесь р₆: «СМИ должны быть независимы и от государства, и от каких-либо организаций», р₇: «необходимо упрочение федерализма», р₈: «одобрение Болонского процесса».

Для анализа рациональности в качестве элементов Σ – непротиворечивого множества логических зависимостей вида $\chi_m \rightarrow \psi_m$ – рассматривались логические связи между относящимися к одним и тем же темам пунктами программ различных партий. Например, для отношения к частной собственности на землю были сформулированы следующие зависимости: $\{J_{1p57} \rightarrow J_{1p88}, J_{1p57} \rightarrow J_{1p84}, J_{1p67} \rightarrow J_{1p88}\}$ и т.п. Здесь р₅₇: «Земля всецело должна быть в государственной собственности», р₆₇: «Продажа земли в частные руки должна быть строго ограниченной», р₈₄: «Земля должна постепенно приватизироваться, но при условии строгого соблюдения законности», р₈₈: «Необходима свободная без ограничений продажа сельскохозяйственных угодий».

В обоих экспериментальных исследованиях для сторонников различных партий были вычислены значения описанных выше критериев рациональности – непротиворечивости δ , близости к «идеальному мнению» ρ , согласованности мнений λ , определено отношение частичного порядка \leq для полученных значений критериев и построена диаграмма этого отношения [30].

Заключение

Формализованный качественный анализ социологических данных расширяет возможности обработки эмпирических социологических данных. Интеллектуальные системы, использующие ДСМ-метод автоматического порождения гипотез,

представляют собой эффективный инструмент такого анализа. Метод, опирающийся на представление социума в виде системы отношений и предоставляющий средства для анализа этой системы и порождения новых отношений, помогает решить две основные проблемы социологии М. Вебера: исследование индивидуального поведения и обнаружение детерминант социального поведения. При этом ДСМ-метод обеспечивает высокую точность предсказания за счет внутренних средств фальсификации [11], полнота же, по-видимому, может достигаться за счет абдуктивной сходимости метода.

Будучи методом интеллектуального анализа данных, способным порождать гипотезы о каузальных зависимостях из БФ и обладающим средствами автоматического расширения и формирования БЗ, ДСМ-метод открывает возможности для развития когнитивной социологии.

Использование формализованного качественного анализа социологических данных способствует объективизации эмпирических социологических исследований, развитию логики и методологии социальных наук. С другой стороны, потребности ФКАСД являются стимулом для развития компьютерной поддержки таких исследований и, стало быть, средств интеллектуальных систем.

Литература

1. Готтлиб А.С. Введение в социологическое исследование (качественный и количественный подходы). М.: Флинта, 2005. 384 с.
2. Татарова Г.Г. Методы анализа данных в социологии. М.: Изд. Дом «Стратегия», 2002.
3. Страусс А., Корбин Дж. Основы качественного исследования. Обоснованная теория. Процедуры и техники. М.: КомКнига. – 2007. 256 с.
4. Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. From Data Mining To Knowledge Discovery: An Overview // In: Advances in Knowledge Discovery and Data Mining, eds. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, AAAI Press/The MIT Press, Menlo Park, CA., 1996, pp. 1-34.
5. Ragin C.C. The Comparative Method: Moving beyond Qualitative and Quantitative Strategies. Berkley, Los Angeles and London: University of California Press, 1987. 185 p.
6. Rihoux B. Qualitative Comparative Analysis and Related Systematic Comparative Methods. International Sociology, v. 21 (5), September 2006.
7. Ragin C.C. Fuzzy-Set Social Science. Chicago: University of Chicago Press, 2000, 384 p.

8. Михеенкова М.А., Финн В.К. Логика интеллектуальных систем как средство системного анализа в социологии // В сб.: Математическое моделирование социальных процессов. М.: Университет. Книжный дом, 2009, вып. 10, с. 236 – 248.
9. Финн В.К. Синтез познавательных процедур и проблема индукции // В кн. Автоматическое порождение гипотез в интеллектуальных системах. Под ред. проф. В.К. Финна. М.: Книжный дом «Либроком», 2009, с. 92 – 158.
10. Михеенкова М.А., Финн В.К. Правдоподобные рассуждения и булева алгебра для анализа социологических данных (проблемы когнитивной социологии) // В сб. Математическое моделирование социальных процессов, М.: Университет. Книжный дом, 2009, вып.10., с. 229 – 236.
11. Арский Ю.М., Финн В.К. Принципы конструирования интеллектуальных систем // Информационные технологии и вычислительные системы, 2008, № 4, с. 4 – 36.
12. Крыштановский А.О. Анализ социологических данных. М.: Издательский дом ГУ ВШЭ, 2006. 282 с.
13. Бурковская Ж.И., Михеенкова М.А., Финн В.К. Об интеллектуальной системе для анализа электорального поведения // IX Национальная конференция с международным участием “Искусственный интеллект-2004”, Тверь, Сентябрь 8-11, 2004, Труды конференции в 2 томах, т.1, с. 120 – 128.
14. Бурковская Ж.И. Архитектура интеллектуальной системы типа ДСМ для формализованного качественного анализа социологических данных // НТИ, сер. 2, 2007, № 1, с.
15. Парсонс Т. О теории и метатеории // В кн.: Теоретическая социология. Антология. Т. 2. М.: Наука, 2002. С. 44–45.
16. Вебер М. Основные понятия социологии // Вебер М. Избранное. М.: РОССПЭН, 2006.
17. Сорокин П. Квантофрения // В кн.: Социология. Хрестоматия для вузов. Составитель А.И. Кравченко. М.: Академический проект, 2002, с. 63 – 74. Оригинал: Quantofrenia //in: Sorokin P. Fads and Foibles in Modern Sociology and Related Sciences. Westport, Connecticut: Greenwood Press Publishers. 1956. P. 102–130.
18. Поппер К. Ницета историцизма. М.: Издательская группа «Прогресс», 1993.
19. Поппер К. Эволюционная эпистемология // В кн.: Эволюционная эпистемология и логика социальных наук. М.: URSS, 2000.
20. Гусакова С.М., Михеенкова М.А., Финн В.К. О логических средствах автоматизированного анализа мнений // В кн. Автоматическое порождение гипотез в интеллектуальных системах, под ред. проф. В.К. Финна. М.: Книжный дом «Либроком», 2009. с. 446 – 484.
21. Михеенкова М.А., Финн В.К. Об одном варианте формализованного качественного анализа социологических данных // X Национальная конференция с международным участием “Искусственный интеллект-2006”, Обнинск, Сентябрь 26-28, 2006, Труды конференции в 3 томах, т.1, с. 284 – 292.
22. Финн В.К. Правдоподобные рассуждения в интеллектуальных системах типа ДСМ // В кн.: ДСМ-метод автоматического порождения гипотез: логические и эпистемологические основания. Под ред. О.М. Аншакова. М.: Книжный дом «Либроком», 2009, с. 10 – 50.
23. Виноградов Д.В. Формализация правдоподобных рассуждений в логике предикатов // В кн.: ДСМ-метод автоматического порождения гипотез: логические и эпистемологические основания. Под ред. О.М. Аншакова. М.: Книжный дом «Либроком», 2009, с. 287 – 293.
24. Luhmann N. Öffentliche Meinung. Politische Planung, Aufsätze zur Soziologie von Politik und Verwaltung. – Opladen. – 1971.
25. Финн В.К., Михеенкова М.А. К формальному определению закрытого социологического опроса // Тезисы докладов III Всероссийской научной конференции Сорокинские чтения: “Социальные процессы в современной России: традиции и инновации” в 5 томах, Москва, 4-5 декабря 2007, М.: Университет. Книжный дом, 2007, Т.1, с. 214 – 217.
26. Климова С.Г. Михеенкова М.А., Панкратов Д.В. ДСМ-метод как метод выявления детерминант социального поведения //В кн. Автоматическое порождение гипотез в интеллектуальных системах, под ред. проф. В.К. Финна. М.: Книжный дом «Либроком», 2009. с. 410 – 427.
27. Финн В.К., Михеенкова М.А. О логических средствах концептуализации анализа мнений // В сб.: Многозначные логики и их применения, т. 2: Логики в системах искусственного интеллекта. Под ред. проф. В.К. Финна. М.: Издательство ЛКИ, 2008, с. 152 – 199.
28. Девятко И.Ф. Социологические теории деятельности и практической рациональности. М.: “АВАНТИ ПЛЮС”, 2003.
29. Финн В.К. Об одном варианте логики аргументации // В сб.: Многозначные логики и их применения, т. 2: Логики в системах искусственного интеллекта. Под ред. проф. В.К. Финна. М.: Издательство ЛКИ, 2008, с. 13 – 58.
30. Михеенкова М.А., Финн В.К. Интеллектуальный анализ данных для проблем когнитивной социологии // XI Национальная конференция с международным участием “Искусственный интеллект-2008”, Дубна, Сентябрь 29 – Октябрь 2, 2008, Труды конференции в 3 томах, т.2, с. 61 – 69.

Михеенкова Мария Анатольевна. Старший научный сотрудник Всероссийского института научной и технической информации РАН (ВИНИТИ РАН). Окончила Московский физико-технический институт (МФТИ) в 1981 году. Кандидат технических наук. Имеет 45 научных работ. Область научных интересов: искусственный интеллект, логика и методология социальных наук, интеллектуальный анализ данных, качественный анализ. mmikh@viniti.ru.