

Г.В. КАНЬГИН

КОМПЬЮТЕРНОЕ АССИСТИРОВАНИЕ В МАССОВЫХ ОПРОСАХ: ПРОБЛЕМНО- ОРИЕНТИРОВАННЫЙ ПОДХОД

Каныгин Геннадий Викторович — кандидат экономических наук, руководитель группы Социологического института РАН. Адрес: 198005 Санкт-Петербург, Измайловский проспект, д. 14. Телефон: (812) 316-21-62; Факс: (812) 316-29-29. Электронная почта: genaka@hotmail.com
Исследование выполнено при поддержке Российского фонда фундаментальных исследований (проект № 00-06-80184) и Российского гуманитарного научного фонда (проект № 00-03-00279а).

ВВЕДЕНИЕ

Разработка систем компьютерного ассистирования интервьюированию (КАИС) связана преимущественно с использованием современных информационных технологий, созданием новых программных средств и другими «общевычислительными» задачами. Получены интересные результаты в области сбора данных [1] и мультимедийного обеспечения обследований [2]. Однако в современных КАИС недостаточно развиты специфические средства ассистирования, и сама проблематика «компьютерного опроса» остается малоизученной. Одним из таких средств является маршрутизация опроса (Automatic Awareness Logic™), реализованная во многих системах. Возможности маршрутизации специально обсуждаются разработчиками системы Tric-Trac [3]. Задачи такого типа возникают именно в процессе опросов, и средства их решения не могут быть заимствованы из другой предметной области. Философия КАИС определяется следующим образом: «Пользователь выполняет работу один раз, но пользуется ее результатами многократно; вводит тексты вопросов и их закрытия только один раз; визуализированные на экране вопросы многократно применяются в документах опроса, таблицах, отчетах и графическом представлении результатов; тексты вопросов используются без повторного ввода; пользователь может создавать шаблоны таблиц для неоднократного применения» [4]. Повторное использование запрограммированных объектов является общим для объектно-ориентированного программирования и не связано с проблематикой массовых обследований.

В основе программных средств Bellview лежит язык спецификации анкеты (QSL) [5]. Разработчики системы ставят своей задачей создание специализированного языка программирования. Но язык программирования и язык опроса не одно и то же. QSL строится по образцам известных языков программирования (например, для разработчиков Blaise образцом служит Pascal [6]) и предназначен для выражения программных объектов: чисел, способов обращения к внешним программам, визуализации и т. п. Тем самым пользователь вынужден изучать терминологию программирования и связанные с ней операции. Конечно, с помощью визуализации, то есть применения визуального языка спецификации анкеты (visual QSL), можно об-

легчить обращение с системой. Однако проблемы собственно анкетирования останутся нерешенными.

В данной статье рассматривается «неинформационный» ресурс систем анкетирования, основанный на анализе структуры анкетного метода как специфического метода социологии. В этом и заключается основная идея проблемно-ориентированной структуры компьютерного ассистирования опроса, разработанной в виде пилотажной версии КАИС.

ИСХОДНЫЙ ТЕЗИС

Анкетный вопрос любого типа (простой, многоальтернативный, метрический, открытый) включает две компоненты. Первая компонента представлена в текстовом виде и выражает сюжет вопроса (СВ) — о чем спрашивают: возраст, доход, научная квалификация и т. п. Вторая компонента — шкала (Ш) — имеет различные формы, обусловленные типом вопроса. Простой, многоальтернативный вопрос — предопределенное множество ответов; открытый вопрос — автоматически обновляемое множество ответов, метрический (количественный) вопрос — описание соответствующей единицы измерения. В любом случае шкала — это предопределенные термины, в которых предполагается формулировка ответа. Таким образом, первая компонента (вводная часть) вопроса представляет собой открытую метафору, то есть вербальную конструкцию, которая по-разному интерпретируется участниками опроса. Вторая компонента выражена предопределенными терминами, которые одинаково понятны всем участникам опроса.

Примеры

Простой вопрос: «Возраст_1»; сюжет вопроса: Сколько лет Вам исполнилось на день опроса? Шкала: 1: 17-24, 2: 25-33...

Метрический вопрос: «Возраст_2»; сюжет вопроса: Каков Ваш возраст на момент опроса? Шкала: год.

Открытый вопрос: «Предметы обихода_1» сюжет вопроса: Перечислите предметы Вашего домашнего обихода; Шкала: ...3: холодильник, 4: телевизор, 5: стиральная машина...

Многоальтернативный вопрос: «Предметы обихода_2» сюжет вопроса: Какие из перечисленных ниже предметов домашнего обихода есть у Вас дома; Шкала: 1: пылесос; 2: микроволновая плита; 3: холодильник...

Важно отметить, что шкала включает не только варианты ответа (1 — мужчина, 2 — женщина), но и единицы измерения (год, рубль, процент); шкала может не быть явно выражена в тексте вопроса, но всегда подразумевается. Предлагаемое членение не имеет исключений по типам вопросов.

МЕТАФОРА В АНКЕТНЫХ ФОРМУЛИРОВКАХ

Следуя М. Блэку, мы будем понимать под метафорой связку между системами общеупотребительных ассоциаций [7]. Открытость метафоры позволяет использовать перенос значений для описания вводной части вопроса. У каждого респондента формируются собственные ассоциации с вводной частью вопроса, то есть собственные «определения» содержащихся в нем лексических единиц. Например, «возраст» может вызвать в принципе бесконечное множество определений. Здесь необходимы кавычки, хотя имеется и «научное» определение возраста. Текст (набор символов, например, в кодировке ANSI) — это изображение стоящей за ним метафоры. Каким образом изображение связано с его метафорой — неизвестно, поскольку модель ме-

тафоры отсутствует. Очевидно, что любое изменение текста означает изменение метафоры. Поэтому «возраст» отличается от «возраста на момент опроса».

Метафора — шкала. Противопоставление «метафора/шкала» является выражением проблемы неоднозначности. При интерпретации вопроса о возрасте один вспоминает детство, другой — предпенсионные хлопоты, третий — проблемы здоровья и т. д. При операционализации шкалы (1 — молодые, 2 — пожилые) на основе единицы измерения «год», можно использовать пороговые значения 33 (возраст Христа), 40 (предельный возраст молодости для российского ученого в терминах фонда Форда). И в том и другом случаях наблюдается явное несоответствие определений, которое в конечном счете приводит к неоднозначности формулировок вопросника. Однако «возраст» остается метафорой (открытой конструкцией), а шкала — однозначно понимаемой количественной конструкцией. Спор «33» и «40» ведется в рамках однозначно понимаемых единицы измерения (год) и отношений порядка. Операционализация, минуя шкалу, равнозначна появлению «духа», поскольку возраст предстает здесь как все «множество» человеческих ассоциаций. Кроме числа прожитых лет, значения возраста можно связать с оборотами Земли вокруг Солнца, внешним видом, наличием детей и внуков. Проблема выбора метафоры решается «практическим разумом» социолога при выборе шкалы.

Любая шкала является метафорой, поскольку в ее градациях присутствуют текстовые изображения и, соответственно, стоящие за ними метафоры, например, (1 — мужчина, 2 — женщина). Но по канонам анкетного метода эти метафоры, во-первых, являются однозначно понимаемыми, во-вторых, их компьютерными моделями оказываются вычислительные операции, выполняемые над кодами (1, 2).

Операция ассоциирования. Противопоставление «метафора/шкала» реализуется операциями с этими объектами. Для метафоры разрешены две основные операции: ассоциирование и редактирование текста изображения. Ассоциирование — это связывание двух метафор, например, «вводная часть вопроса — вводная часть вопроса», «шкала — шкала» и т. д. Для выражения однозначности используется ассоциирование «вводная часть вопроса — шкала». В отличие от метафор, для шкал наряду с ассоциированием и редактированием текста доступны вычислительные операции. Эти операции могут быть дополнительно типологизированы, например, в терминах социологического измерения. Так, возможно разделение шкал на номинальные, ранговые и метрические.

ПРИНЦИПЫ МОДЕЛИРОВАНИЯ СТРУКТУРЫ АНКЕТНОГО МЕТОДА

Чтобы проследить исходное членение анкетного вопроса на всех стадиях анкетирования (разработка анкеты, интервьюирование, создание индексов, обработка данных, интерпретация результатов, администрирование данных) необходимо соблюдать следующие принципы.

Ассоциативность. В анкетном методе получает структурное выражение фундаментальная проблема соотношения количественных и качественных подходов в социологическом исследовании. Вводная часть вопроса, являя собой открытую метафору, передает качественную сторону социологи-

ческого анализа. Шкала — измерительная процедура — выражаемая в математических или естественнонаучных терминах и одинаково понимаемая в соответствии с канонами анкетного метода, воплощает количественную традицию. Принцип ассоциативности заключается в том, что формулирование однозначно понимаемых вопросов сводится к решению эвристической задачи связывания в единой вопросной конструкции контекстно зависимого описания (открытой метафоры) и контекстно свободного описания (шкалы). Такое связывание не может быть выполнено с помощью каких бы то ни было количественных методов. Принцип ассоциативности отражает требование однозначности формулировок вопросов при анкетировании, показывая, каким образом количественные конструкции вводятся в рамки метода.

Парадигмальность. Вводная часть вопроса (вводная метафора) по своей природе является контекстно зависимой, то есть такой конструкцией, понимание которой невозможно без знания сюжета анкетирования. Поэтому принцип ассоциативности не указывает правил связывания вводных метафор и шкал, предоставляя здесь исследователю полную свободу. Однако свобода связывания возможна только до тех пор, пока один исследователь в соответствии с принципом ассоциативности не ограничивает свободу другого исследователя, осуществляющего тот же принцип. Поэтому ни одна из реализаций связывания (ни один из вопросов-ассоциаций), созданных разработчиком, не должны противоречить аналогичным ассоциациям, практикуемым другим разработчиком.

Метафоризированная логика. Любая логическая или математическая операция имеет смысл только в том случае, если ее составляющие (термы, выражения, предложения и т. п.) понимаются одинаково. В соответствии с принципом ассоциативности это означает, что различные вычислительные (логические, математические, формальные) операции, характерные для анкетирования, проводятся над шкалами, а не над вводными частями вопросов.

Порождения. Множество значений (числовых, текстовых, графических и т. д.), с которыми работает метафоризованная логика, появляется в два этапа: метафора - шкала, шкала - значение шкалы. Для реализации этих переходов следует учитывать, что механизм метафоротворчества в отличие от шкалы, не моделируется компьютером.

ПРОБЛЕМНО-ОРИЕНТИРОВАННЫЕ МЕХАНИЗМЫ АНКЕТНОГО МЕТОДА

Разработка КАИС на основе выдвинутых принципов требует пересмотра понятий анкетного вопроса, блока вопросов, шкалы, анкеты, данных опроса и т. д. Рассмотрим два новых понятия — понятие парадигмального сходства и понятие вычислительных шкал, которые представляют собой оригинальные средства компьютерного ассистирования.

Парадигмальное сходство. Существование «нормальной науки» (в смысле Т. Куна) предполагает существование исследовательского коллектива — носителя парадигмы или дисциплинарной матрицы значений. В социологической науке это находит выражение в устойчивой традиции применения анкетного метода и накоплении результатов исследований (в том числе в машинном виде). В различных исследованиях одна и та же вводная часть вопроса может ассоциироваться с разными шкалами и наоборот: одна и та же шкала может связываться с различными вводными метафорами. Таким

образом, оказывается, что вводные метафоры (как и шкалы) могут быть сходными не в смысле некоторой гипотетической меры близости, а в связи с операцией ассоциирования. Последняя отражает опыт анкетных обследований в рамках направления, практикующего анкетный метод, и не предполагает универсалий типа признакового пространства или системы координат.

Сюжеты вопросов, ассоциированные в различных исследованиях с одной и той же шкалой, называются парадигмально сходными. Шкалы, ассоциированные в различных исследованиях с одной и той же вводной частью, также называются парадигмально сходными. Развиваемый подход позволяет использовать компьютерно-организованные мнемонические средства.

Примеры парадигмального сходства

Разработка анкеты. Для сюжетов вопросов: СВ1: «Сколько Вам лет» и СВ2 «Укажите Ваш возраст» исследователь в разных опросах указал одну и ту же шкалу: ...1: 17-24, 2: 25-33... Следовательно, СВ1 и СВ2 являются парадигмально сходными.

Для одного и того же сюжета вопроса «Укажите Ваш возраст» в разных исследованиях применены две шкалы — Ш1 в годах, Ш2 ...1: 17-24, 2: 25-33... И в этом случае сюжеты вопросов являются парадигмально сходными.

Обработка данных. В одном из исследований задание на обработку составлено для СВ1 «Среднемесячный доход». При использовании этого же задания для обработки результатов другого исследования оказывается, что СВ1 не представлено в данных, зато имеется СВ2 «Душевой доход в семье». Если исследователь при повторном счете укажет СВ2 для замены в задании СВ1, то тем самым СВ1 и СВ2 станут парадигмально сходными.

Использование парадигмального сходства. Текст — это изображение стоящей за ним метафоры и любое изменение текста может вести к ее подмене. Ясно, что во многих случаях изменение текста не ведет к появлению новой метафоры. Например, вопросы «Каков Ваш возраст?», «Сколько Вам лет?» и «Сколько Вам лет на текущий момент?» тематически ориентируют респондента на одну и ту же метафору его возраста, хотя первое текстовое изображение отличается от второго и третьего.

Для решения вопроса о том, насколько меняется метафора при редактировании текста, создается тот или иной алгоритм, в рамках которого по формальным признакам сходства текстов выносится решение об идентичности стоящих за ними метафор. Как правило, это решение неокончательное и дополняется дальнейшим установлением «истины». Например, для приведенного выше примера алгоритм сходства второго и третьего текстов может сводиться к правилу: если три слова идентичны, то метафора не изменилась.

Идея заключается в том, чтобы максимально структурировать работу при анкетировании и с помощью КАИ поставить участников опроса в такие условия, при которых (а) удобно ассоциировать тексты, появляющиеся в анкетном методе; (б) связывать их с predeterminedными терминами; (с) накапливать результаты ассоциирования и связывания в компьютерной системе. Тогда КАИ «автоматически» определяет сходство текстов не только в связи с их формальной идентичностью, но и в «человеческих» смыслах. В этом случае мы не сможем пропустить сходства конструкций «Каков Ваш

возраст?» и «Сколько Вам лет?». В работе проблемно-ориентированной КАИС принуждение социолога к ассоциированию метафор и шкал предстает как компьютерное ассистирование, а человеческая непоследовательность в формулировках текстов — как способ накопления знаний.

Таким образом, вводные метафоры (равно как и шкалы) могут быть сходными в связи с операцией ассоциирования; последняя отражает опыт анкетных обследований в рамках исследовательской парадигмы и не предполагает универсалий типа признакового пространства или системы координат; парадигмальное сходство принимает различную форму (в частности, в зависимости от фазы опроса), вводные части вопросов могут быть сходными в рамках обработки данных, разработки анкеты, построения новых индексов и т. д., аналогичное утверждение справедливо и для шкал; парадигмальное сходство переводит проблему семантического сходства текстов в структурирование и организацию использования КАИС.

Вычислительные шкалы. В соответствии с принципом метафоризованной логики все вычислительные задачи в анкетном методе реализуются на основе представления о шкале. К таким задачам принадлежат: построение новых признаков (сюда же относятся операции по моделированию выборочного метода — квотные задания, автоматизация их контроля и т. п.), обработка данных, маршрутизация, фильтрация данных, Automatic Awareness Logic™¹.

Обращение к вычислительным средствам, решающим эти задачи, обусловлено, главным образом, традициями соответствующей предметной области. Соответствующие программные средства представлены, например, командными языками в статистических пакетах, формулами фильтрации и языками запросов в СУБД, алгоритмами поддержки квотного задания и маршрутизации в КАИС и т. п. В то же время понятие шкалы, с точки зрения анкетного метода, канонизировано как набор вариантов ответов. Поэтому представление о шкале должно быть сконструировано таким образом, чтобы, с одной стороны, эта конструкция оставалась легко узнаваемой конечным пользователем, с другой — позволила обращаться к средствам компьютерного ассистирования.

При разработке модели вычисляющей шкалы необходимо учитывать ее включенность в операцию ассоциирования. Решить данную проблему можно с помощью вычисляющей шкалы (ВШ). Не вдаваясь в детали программной реализации, укажем ее особенности. Наряду с вариантами ответов шкала описывает единицы измерения, в терминах которых предполагается ответ; тексты градаций шкалы могут включать командный язык, который (а) позволяет формировать процедуры решения характерных вычислительных задач (метризация, квантификация, перекодировка, фильтрация данных, маршрутизация, создание динамических блоков вопросов) на любой фазе анкетирования; (б) имеет для всех фаз анкетирования один и тот же пользовательский интерфейс; (в) допускает обращение к внешним вычислительным средствам (DLL).

Использование вычислительных шкал

Построение новых индексов. Социологу необходимо узнать, сколько молодых и пожилых людей он опросил. В данных присутствуют результаты опроса со шкалой «год». Создается новый индекс («квазивопрос»), который

представляет собой, во-первых, вводную метафору, стоящую, например, за словами «квантифицированный возраст», во-вторых, вычисляющую шкалу, в которой в данном случае используется исходная шкала «год», правила квантификации (1: ≤ 33 , 2: > 33), градации (1 — молодые, 2 — пожилые). Получаемый таким образом «квазивопрос» в проблемном плане не отличается от исходного вопроса о возрасте, так как «квантифицированный возраст» — это неоднозначная метафора, а привязываемая к ней шкала (1 — молодые, 2 — пожилые) является однозначно понимаемой конструкцией, поскольку определяется через однозначно понимаемую единицу измерения «год» и правила квантификации, обладающими тем же свойством однозначности. В данном случае уже не имеет значения, что слова «молодые, пожилые» являются метафорами, за которыми у каждого из респондентов стоят личные ассоциации. Главное, что определение понимается однозначно. Фактически однозначно интерпретируемая шкала формируется при ответах респондентов в соответствии с исходной шкалой. При создании вычисляющей шкалы пользователю, конечно, придется указать исходную шкалу и выполняемые операции, что отличит вычисляющую шкалу от традиционной шкалы как множества вариантов ответов. Однако в процессе использования такой шкалы операции по ее ассоциированию с сюжетом квазивопроса ничем не отличаются от таковых для обычных вопросов респонденту. При накоплении результатов опросов, получаемых с помощью проблемно-ориентированной КАИС, относительные трудозатраты по созданию шкал уменьшаются, так как возникающие наработки облегчают работу по формулированию новых количественных выражений. Для вычисления индекса (не путать с фазой создания) используется аппарат парадигмального сходства, то есть в случае отсутствия исходной шкалы среди представленных в данных вопросах проводится автоматический поиск и предъявление парадигмально сходных шкал.

Интервьюирование. При интервьюировании, например, в модуле самоопроса респондента, вычислительные шкалы используются для двух целей: изменения порядка предъявления вопросов (ветвления процесса опроса) и динамического изменения состава вопросов в блоке. И то, и другое зависит от ответов респондента. Первая задача известна как маршрутизация, вторая описана на сайте Tric-Trac [3] под названием Automatic Awareness Logic™. Наш подход позволяет организовать решение обеих задач одним и тем же вычислительным аппаратом и с помощью единого пользовательского интерфейса. Кроме того, оба случая могут трактоваться как расширение представления об ассоциировании компонентов вопросов. Маршрутизация автоматически выполняет операцию ассоциирования, сопоставляя вариант ответа респондента и следующий вопрос интервью. Хотя такого рода переход планируется при разработке анкеты, но до опроса остаются неясными количественные параметры ассоциирования. Например, какие именно маршруты представлены и как часто опрос проводится в соответствии с ними. Динамический блок связывает группы вопросов, которые также могут быть интервьюированы как парадигмально сходные. Тем самым оказывается, что вычислительная шкала расширяет область применения анкетного метода до всех ситуаций, где может быть использован механизм парадигмального сходства.

Таким образом, одно и то же понятие шкалы используется и с целью фиксации вариантов ответов, и с целью количественных преобразований; формирование вычислительной шкалы для всех вариантов ее использования обеспечено единым пользовательским интерфейсом; применение вычислительной шкалы расширяет круг ситуаций, в которых используется ассоциирование вопросов и их компонентов, то есть развивает механизм парадигмального сходства.

ОБСУЖДЕНИЕ

Тотальная структуризация КАИС. Путем структуризации вопроса на две компоненты, поддержания отношения «метафора/шкала» на всех фазах интервьюирования и «автоматического принуждения» социолога выбирать конкретное сочетание метафоры и шкалы, система становится способной «вытаскивать» из исследователя его эвристическое знание о том, как отразить сюжет вопроса в понятных терминах. Без «тотальной» структуризации всех компьютерных инструментов КАИС достичь этой цели невозможно. Одним из механизмов решения задачи является парадигмальное сходство. Что касается вычислительных шкал, то в данном изложении они представляют собой инструмент единого пользовательского интерфейса. Однако в программной реализации вычислительные шкалы являются объектами, с которыми связаны различные проблемно-ориентированные события, что заметно расширяет функциональные возможности этого инструмента.

Эвристические возможности. В составе любой КАИС присутствуют как методы собственно анкетного ассистирования, так и общенаучные средства типа статистической обработки или выборки. Те и другие методы основаны на шкалах, то есть могут быть введены в действие только после решения социологом эвристических задач. Мы предлагаем инструмент, обеспечивающий ассистирование на ранней фазе опроса, когда комбинации «метафора/шкала» находятся в стадии создания.

Человеко-машинный принцип. Обычно вычислительный процесс (например, технологический поток) на каком-то этапе прерывается и человеку предлагается оценить значения некоторых параметров. Он выполняет необходимые оценки, значения которых возвращаются вычислительному процессу, и он возобновляется. «Человеческое» решение — это этап вычислительного процесса, то есть человек выполняет решение частной задачи, с которой справляется лучше машины, а общая задача остается в ведении компьютера. В нашем случае человек лучше справляется с общей задачей (создает метафоры), а вычислительный процесс (накопление их текстовых изображений, графическое представление результатов, построение новых признаков) — это частная задача и ее лучше решает машина.

Указанные принципы — попытка создания архитектуры вычислительной системы в условиях, когда неизвестны способы моделирования ключевого механизма (метафоротворчества), для ассистирования работе которого создается сама система. Естественно, реализация принципов предполагает человеко-машинную организацию ассистирования.

Еще одной особенностью предложенного человеко-машинного подхода является то, что человек не является «единым» экспертом. Социолог используется как эксперт по выбору количественных представлений, в рамках ко-

торых респонденты как эксперты по интересующей социолога проблеме выбирают значения ответов. В результате такой двойной экспертизы и становятся возможными компьютерные выкладки.

ЗАКЛЮЧЕНИЕ

Особенности проблемно-ориентированного подхода к компьютерному ассистированию опроса заключаются в следующем. Ключевая проблема анкетирования — создание однозначных анкетных формулировок — получает здесь структурное выражение. Предлагается механизм порождения сходства объектов, изучаемых с помощью анкетного метода, который принципиально отличен от подходов, построенных на универсалиях типа признакового пространства или характеристического множества. Кроме известных приемов компьютерного ассистирования (маршрутизации, Automatic Awareness Logic™), являющихся по своей природе логическими, социологу становятся доступными эвристические средства, выражающие его исследовательскую интуицию. Подход воплощает современные тенденции объектно-ориентированного программирования, значительно сокращает объемы работ по написанию КАИС и создает условия для ее дальнейшего развития.

ЛИТЕРАТУРА

1. *Leeuw E., Nicholls W.* Technological innovations in data collection: Acceptance, data quality и costs // Sociological Research Online. Vol. 1. No. 4. 1996. <<http://www.socresonline.org.uk/socresonline/1/4/leeuw.html>>
2. *Macer T.* Article #35: ASC99 Conference report. 1999. <<http://www.macer.co.uk>>
3. TRIC TRAC: <<http://www.trictrac.com/product.htm>>
4. Sphinx: <<http://www.scolari.co.uk/>>
5. Bellview: <www.ultersys.msk.su/bvcati.htm>
6. Blaise: >www.westat.com/blaise/blaisefaq.html>
7. *Black M.* Models and metaphor: Studies in language and philosophy. Ithaca-London: Cornell University Press, 1962.

ПРИМЕЧАНИЕ

1. Создание SQL-запросов не включается нами в понятие шкалы, так как трактовка данных в терминах SQL значительно отличается от их трактовки при анализе результатов опросов.