

4. МЕТОДЫ ИССЛЕДОВАНИЯ РЫНКОВ

Тенденции развития маркетинговых и социальных исследований

Демидов Александр Михайлович,
*Международный институт маркетинговых
и социальных исследований «ГфК-Русь»*

Вчера

Прежде чем говорить о современных тенденциях практически ориентированных исследований, к которым относятся маркетинговые, медиа- и прикладные социальные исследования, мне бы хотелось взглянуть на состояние российской индустрии исследований 10 лет назад, на рубеже 21-го века.

Одним из важнейших изменений в технологии сбора данных в тот период стало внедрение в практику крупнейших исследовательских компаний систем автоматизированного компьютеризированного телефонного интервью или CATI (Computer Assisted Telephone Interview), которые быстро потеснили личные интервью. До этого телефонные опросы мало применялись, как нерепрезентативные в связи с недостаточной телефонизацией населения и к тому же они мало отличались по себестоимости от личных интервью. Внедрение CATI привело к экономии бюджетов исследования и стало особенно эффективно для массовых опросов, поскольку экономия средств на каждом телефонном интервью значительно превышает дополнительные затраты на специальное оборудование и программирование анкет. В настоящее время специально оборудованные CATI студии имеют практически все крупные и средние исследовательские компании и организации.

Другой инновацией в начале нулевых годов 21 века стало внедрение опросов с использованием компьютеров, CAPI (Computer Assisted Personal Interview) опросов. Это также был большой шаг в развитии современных технологий исследований, позволивший существенно облегчить труд интервьюеров и внедрить интерактивные методики опроса респондентов. Одним из ключевых факторов успеха технологии CAPI стала мультимедийность: возможность собирать аудио- и видеоматериалы.

В полную силу на рубеже 21-го века заработали панельные исследования: медийная, потребительская и торговая панели. Правда, за исключением телевизионных медиаизмерений, в панелях использовались дневниковые, бумажные методы сбора данных и физическая регистрация. Хотя российские исследователи знали об использовании на Западе технологий онлайн исследований (CAWI — Computer Assisted Web Interview), распространения они не получали из-за низкой пенетрации Интернета. В содержательном плане на рынке появились специализированные исследования:

- бренда и имиджа,
- коммуникаций,
- ценовые,
- лояльности и удовлетворенности и т.д.

В первую очередь это было связано с взрослением российского потребительского рынка, когда главной бизнес-проблемой становится не задача доставки товаров потребителю, а задача обаять и привязать потребителя к товару/услуге.

Растущее понимание того, что уловить тенденции на рынке и в обществе можно только в динамике, привели к широкому распространению трекинговых и мониторинговых исследований.

Растущая потребность рынка не только понимать, но и уметь управлять поведением потребителей привели к внедрению в практику исследований, наряду с пассивными методами измерений (опрос, наблюдение, регистрация), также и активных методов (эксперимент). Наиболее

яркими представителями экспериментальных методов являются методы тестирования продуктов и совместный анализ (conjoint).

На рубеже 20-го века сформировался рынок профессиональных интервьюеров, превратившийся в самостоятельную профессию.

Покупкой TNS компаний MIC и Gallup Media завершился выход на российский рынок всех основных мировых игроков маркетинговых исследований, которые в этот период заняли лидирующее положение и в России.

Сегодня

Сегодня рынок маркетинговых, медиа- и прикладных социальных исследований вступает в новую фазу своего развития, которое, в свою очередь, определяется бурным развитием новых цифровых технологий. В первую очередь это касается онлайн-исследований.

За прошедшие 10 лет Интернет стал неотъемлемой частью нашей жизни. Постепенно эта новая медиа-среда становится основной платформой маркетинговых исследований: и как способ сбора данных, и как самостоятельный объект для исследования.

Сегодня большинство крупных исследовательских компаний имеют свои онлайн панели, причем, будучи сформированным в отдельных странах, эти панели интегрировалась в глобальную панель. Например, размер панели всей ГфК Группы составляет 1 500 000 (полтора миллиона) респондентов, ГфК–Русь — 65 000 респондентов, и они постоянно пополняются. Сегодня существуют крупные компании, которые специализируются только на создание, поддержание панели и проведении опросов, например, компания OIM. Налицо явная тенденция увеличения доли онлайн опросов. К этому подталкивают и запросы клиентов. Во-первых, для клиента все более важны быстрота проведения опроса. Во-вторых, клиенты придают все меньшее значение репрезентативности выборки. Для них, прежде всего, важны люди, соприкасающиеся с их продуктом, т.е. целевые группы. А онлайн панель, будучи столь масштабной и структурированной позволяет выделять практически любые целевые группы.

Интернет становится мощным фактором формирования потребительского поведения. Огромное количество ценнейшей для исследователя информации содержится в социальных сетях, блогах, на сайтах, форумах. Все более распространение получают панели респондентов, компьютеры которых оборудованы специальным софтом, который отслеживает и фиксирует все их поведение в Интернете.

В русле этих тенденций компания ГфК–Русь в этом году запускает технологии трекинга поведения пользователей в Интернете (регистрация посещений веб-ресурсов, контакты с рекламой) и метод онлайн-опроса для выявления эффектов рекламы (например, узнаваемость бренда, его оценка, готовность к покупке и пр.).

Одним из направлений сегодня является изучение удобства интерфейсов, так называемое usability и наблюдение за поведением людей в процессе использования софта.

В исследовательских целях уже используются, хотя и далеко не полностью, возможности мобильных телефонов и мобильного Интернета.

В России уже активно применяется метод «data mining» для добычи информации из Интернета, например, из тех же социальных сетей.

Если говорить о других направлениях в развитии исследовательских технологий, то в настоящее время уже широко применяются нейрофизиологические методы, например, Eye-tracking с дополнительным снятием биологических реакций респондента. В результате таких исследований мы получаем данные о концентрации внимания исследуемого и его эмоциональном состоянии. Как показывает мировая практика, исследования эмоций получают все большее распространение, так как обнаруживается их огромная роль в процессах принятия решений, в том числе при покупках товаров и услуг.

Мы наблюдаем более широкое использование наблюдения как метода исследования: например, этнографические исследования и Mystery shopping («таинственный покупатель»).

Для решения ряда исследовательских задач используются методы, позволяющие получить различную информацию из одного источника (single source surveys). Это, например, потребительская панель домохозяйств с одновременным мониторингом просмотра Интернет-рекламы.

В контексте внедрения новых технологий необходимо упомянуть, что потребительская панель домохозяйств в России уже переведена с бумажных дневников на сканеры, снимающие данные о купленных продуктах со штрих-кодов и автоматически передающие собранную информацию в центр обработки и анализа данных.

Внедрение перечисленных технологий начало достаточно быстро менять рынок маркетинговых исследований и его структуру. Чтобы не отстать от конкурентов, все исследовательские агентства просто обязаны ориентироваться на новые технологии как в сфере сбора и анализа данных, так и организации производства. Многие крупные компании покупают фирмы, специализирующиеся на развитии технологий. Небольшие агентства ищут свои технологические ниши. Сбор данных с помощью бумажных анкет неуклонно снижается. На смену «бумажному» полю приходят технологические операции, которые включают в себя одновременно сбор данных, обработку и анализ.

Изменяется и формат поставки данных клиентам. Традиционные многостраничные печатные отчеты сменяются электронной передачей данных и непосредственным доступом клиентов к базам данных агентства. Для клиентов все более важным становится тайминг и консалтинг. Они требуют большей включенности и понимания исследователем проблем своего бизнеса, а не просто распределения ответов на те или иные вопросы. В этой связи в агентствах на первый план выходит не *research*, а *client service*. В результате агентства переходят на новую структуру своей деятельности, включающую два основных направления: *Client Service* и *Operations*.

В значительной степени изменяются принципы сбора данных и роль интервьюеров. Агентства и клиенты все больше общаются непосредственно с респондентами или сами наблюдают за их поведением (этнография, *mystery shopping*), появляются все новые и новые разновидности панелей респондентов.

При этом доступ к респондентам становится все более сложным. Одной из причин является увеличивающаяся мобильность населения: миграционная, туристическая, виртуальная и т.д. Другая причина сложности достижения респондентов — изменение отношения к своему времени, как самостоятельной ценности. Третья причина — невысокая ценность маркетинговых исследований в глазах населения. В этой связи все больше личных опросов проводится в местах скопления публики: торговых точках, кафе, на автостоянках и т.д.

Завтра

Попробуем теперь заглянуть в будущее. Не вызывает сомнений, что в ближайшие 5–10 лет все интернет и мобильные технологии, оказывающие решающее влияние на рынок маркетинговых исследований, получат дальнейшее развитие. Это предполагает, с одной стороны, развитие широких возможностей для исследовательской индустрии, а с другой — большие трудности. Озвучивается, например, такое крайнее мнение, что традиционные маркетинговые исследования вообще отомрут в связи с обилием информации в Интернете. Действительно уже сегодня все больше информации в открытом доступе. Говорят и о том, что с созданием искусственного интеллекта исчезнет необходимость опрашивать людей.

В то же время большинство экспертов считает, что маркетинговые исследования приспособятся к достижениям технологической революции и используют их в своих целях.

Прежде всего, укрепится взаимодействие с мобильными и Интернет провайдерами и операторами социальных сетей, что даст исследователям огромные возможности новых способа сбора данных.

Появятся приспособления, которые позволят считывать мысли, что, наряду с этическими проблемами, может создать дополнительные возможности для исследователя.

Получат дальнейшее развитие нейронаучные методы исследований. Респонденты будут носить приборы, измеряющие биологические реакции. В результате появятся панели профессиональных «окольцованных» респондентов, поведение которых будет мониториться. Они будут называться не респондентами, а участниками исследования.

Клиенты будут стремиться к непосредственному общению с респондентами, что поставит вопрос об изменении роли исследовательских агентств.

В исследованиях будут использоваться приборы, определяющие, говорит ли человек правду.

Изменится персональный состав исследовательских агентств. Все более востребованными будут IT специалисты. В связи с тем, что клиенты все больше будут стремиться к самостоятельному сбору данных, будет происходить «утечка» мозгов из агентств на сторону клиента.

Все больше исследователей (и исследований) будет базироваться в Азии.

С большой долей уверенности можно предположить, что значение и доля опросов в маркетинговых исследованиях будут падать, а сбор данных с помощью наблюдения и технологической регистрации расти. В этой связи количественные и качественные исследования будут сближаться и осуществляться в одних и тех же подразделениях.

Вполне вероятно, что будут созданы гигантские онлайн панели, которые совместно будут использовать различные исследовательские агентства.

Несмотря ни на что, большинство экспертов сходятся во мнении, что face-to-face и телефонные опросы все же сохранятся для решения многих маркетинговых задач.

Роль методов self-ethnography в тестировании концепции нового бренда

Богомолова Людмила Николаевна

Высшая школа менеджмента СПбГУ

Различные виды тестирований занимают на сегодняшний день важнейшее место в индустрии маркетинговых исследований. Потребность в изучении потребительских реакций на продукты и коммуникативные конструкции обусловлена особенностями принятия управленческих решений в индустриях с высоким весом маркетинговой составляющей (прежде всего, FMCG).

Наиболее сложным этапом тестирования нового продукта является тест концепции («concept test» в традиционном исследовательском глоссарии). Этот этап является единственным, на котором ключевую роль играют не количественные, а качественные методы исследования. Сложившаяся исследовательская практика, а также классические работы в области исследований рынка отвергают использования качественных методов в силу их неспособности дать статистически надежный выбор из нескольких объектов. Низкий уровень формализации делает результаты качественного тестирования крайне уязвимыми с точки зрения финального анализа.

Тестирование концепций предполагает иной подход, и в этом случае, специфика качественных методов позволяет добиться значимых результатов. Это связано в первую очередь с природой самого феномена креативной концепции, которая уже более 30 лет активно используется при создании новых продуктов, брендов и, в последнее время, медийных продуктов.

Креативная концепция для нового бренда (или продукта) по форме представляет собой текстовое сообщение, которое кратко отражает все элементы архитектуры бренда, сообщает потребителю о его уникальности и всех тех выгодах, которые он несет в себе. С точки зрения содержания, потребителю предлагается квинтэссенция маркетинговой идеи без возможности обоснованной критики: тестируемый концепт не содержит очевидных недостатков, не имеет физического воплощения, таким образом, потребитель не может давать обоснованные оценки функциональных недостатков продукта. Эта специфика предполагает высокий уровень абстракции всего исследовательского процесса. Соответственно, тестирование не может основываться на идее моделирования ситуации реального выбора (к чему стремятся все виды количественных тестирований различных элементов маркетингового микса).

Наиболее распространенными методами тестирования креативных концепций, на сегодняшний момент, являются фокус-группы (различной структуры и степени формализации) и глубинные фокусированные интервью¹. В современной практике чаще всего используется метод фокус-групп в силу его относительной дешевизны и очевидной экономии времени. Групповая динамика, использование которой составляет ключевую особенность метода, далеко не всегда способствует раскрытию креативного потенциала участников, но имеет значительную ценность с точки зрения побуждения респондентов привлекать собственный социальный опыт для оценки тестируемых объектов.

¹ Мельникова О.Т. Фокус-группы: методы, методология, модерирование: Учебное пособие для вузов. М.: Аспект Пресс, 2007. С. 46.

Качественное тестирование строится на многофакторном анализе всех когнитивных и эмоциональных уровней концепции. Необходимо отметить, что, несмотря на линейный характер внешней структуры исследования (последовательная проверка всех элементов концепции), фактически исследователь работает одновременно на различных уровнях, задавая вопросы, ответы на которые затрагивают все важные для респондента моменты¹.

Если использовать терминологию теории рекламы, то первый уровень, требующий проверки можно назвать когнитивным; необходимо проверить способность концепции к донесению информации. Оценивание этой способности затруднено тем, что предложенные для тестирования материалы могут (и должны) неоднократно изменяться в процессе развития продукта. На этом уровне метод вполне справляется, поскольку происходит фиксация спонтанных оценок понимания информации, которые, как правило, носят достаточно четкий характер.

Наиболее значимые уровни восприятия — ценностные. Специфическая задача исследователя — определить релевантность заложенных в концепцию ценностей, как с точки зрения потребностей целевой группы, так и с точки зрения ценностей продуктовой категории. Обеспечение релевантности является залогом того, что бренд сможет адекватно доносить до потребителей свои ценности и существовать в согласии с их личными потребностями. Соответственно, заказчики исследования уделяют особое внимание диагностике ценностной платформы бренда.

Следующий уровень — эмоциональность, которая также должна соответствовать потребностям целевой группы, позиционировать продукт как уникальный и позволять работать на требуемых ценностно-эмоциональных территориях. Эмоциональность обеспечивает энергетику бренда, а также стимулирует вовлеченность потребителя.

Необходимо учитывать, что для подавляющего количества новых брендов наиболее значимы именно ценностно-эмоциональные выгоды, а для ряда индустрий они стали единственно значимыми. Функциональные преимущества тестируемого объекта или находятся на периферии сознания или связаны с эмоциональными преимуществами. Изучение этих аспектов — традиционная территория качественных методов, поскольку предполагает детализацию потребительского мнения, свободный характер дискуссии и активную роль исследователя в процессе тестирования. Значительным преимуществом качественного подхода в данном случае является и возможность рефлексии исследователя относительно социокультурного обоснования реакций респондентов.

Однако в современной практике тестирования концепций, крайне мало внимания уделяется невербальной составляющей групповой работы: исследователь сосредоточен на вербальном поведении группы и стремится зафиксировать и объяснить именно его.

Традиционный дизайн тестирования концепции предполагает обязательное изучение следующих параметров: понимание содержания; общая эмоциональная привлекательность или непривлекательность; понимание и оценка ключевой идеи; уровень релевантности потребностям целевой аудитории (или релевантности уже имеющейся идентичности бренда, в случае, если тестируется расширение бренда или новая модель и пр.), этическая окраска концепции; уникальность. Этот список не носит иерархического характера и может расширяться в зависимости от специфики исследования.

Таким образом, качественное тестирование предполагает, прежде всего, изучение ценностной релевантности и эмоциональной привлекательности концепции. Традиционно наиболее продуктивен здесь феноменологический анализ в сочетании с этнометодологическими инструментами. Опора на символические модели сознания позволяет провести критический анализ восприятия тестируемых концепций с достаточной точностью для формулировки исследовательских рекомендаций.

Несмотря на высокую продуктивность качественных методов, исследователи постоянно испытывают трудности при разработке дизайна исследования и при анализе его результатов. Некоторые трудности можно отнести к методологическим, другие связаны с предъявляемыми заказчиками требованиями к результатам.

¹ Puri A. The Web of Insights: the art and practice of webnography // International Journal of Market Research. 2007. № 4. P. 49.

Так, в процессе тестирования постоянно возникает необходимость глубокого анализа значительного количества объектов, с высокой степенью схожести (в процессе исследования может тестироваться до 20 концепций; часто они объединены в тематические группы, состав которых может оптимизироваться и изменяться в ходе исследования).

Далее, кроме получения спонтанных реакций по всем значимым критериям, необходимо выяснить базовые ценности и эстетические требования, которые и определяют оценки респондентов. Именно на этом этапе исследования появляется почва для потенциальных исследовательских ошибок и разногласий. Это провоцируется резким дефицитом времени и воздействием групповой динамики, которые предполагают высокий темп задавания вопросов и, в ряде случаев, не дают исследователю возможности для тщательного анализа ответов и обязательного зондирования потребителей.

Очевидна необходимость расширить исследовательский инструментарий. Однако остается актуальной проблема субъективизма в оценках исследователя, которая порождается спецификой метода (и отчасти, лежит в его основе), что крайне нежелательно для формулирования рекомендаций клиенту, особенно на начальном этапе исследования, который чреват возможностью фундаментальных ошибок в построении бренда.

Качественная методология позволяет гибко комбинировать различные исследовательские подходы в рамках одного проекта. Использование методов этнографических исследований (наблюдение, анализ текста, сбор артефактов) приобрели высокую популярность в исследованиях, прежде всего, обращенных к жизненному миру потребителей (изучение моделей поведения, механизмы принятия решений и потребительские практики). Этот же метод, адаптированный под потребности изучения концепций и коммуникаций значительно повышает валидность результатов тестирования.

Наиболее инструментальным является метод, так называемой *self-ethnography*, т.е. этнографические исследования, подразумевающие самостоятельный отбор респондентом артефактов и элементов своего образа жизни, представляемых для исследования. Этот метод лишен спонтанности, которая сопутствует традиционной этнографии, но взамен позволяет изучить логику отбора элементов исследования, проследить, на что опирается потребитель, вынося суждения о ценностях, эмоциях и их воплощении в концепции нового бренда.

В большинстве случаев используется ведение респондентами тематических дневников (блогов) в интернете и/или участие в виртуальных сообществах, которые организуются за некоторое время до начала фокус-групп (интервью) проводимых в режиме реального общения. Тематические дневники охватывают стиль жизни респондента, который предстает в мультимедийной форме (респондент иллюстрирует свои высказывания ссылками на изображения, различные обсуждения, художественные тексты), что позволяет оценить его релевантность как источника оценок и информации. В процессе ведения дневников, респонденты отвечают на вопросы, которые имеют прямое отношение к теме будущего группового обсуждения: это может быть анализ важнейших морально-этических установок; рассмотрение опыта использования различных категорий и пр. Крайне важно задавать те вопросы, которые смогут впоследствии стать платформой для высказываний уже по поводу тестируемых объектов.

Кроме того, модерлируемое сообщество позволяет проводить скрининг коммуникативных особенностей респондентов и их предварительную сегментацию, что имеет особое значение при тестировании креативных концепций. Уровень сложности тестируемых материалов достаточно часто предполагает особые требования к участникам проекта, которые могут быть реализованы только в режиме длительного общения с исследовательской командой.

Наиболее высокий уровень качества предварительного этапа достигается в ситуации, когда исследователь заранее располагает готовыми для тестирования образцами креативных концепций. Это позволяет формулировать задания с учетом тех ценностей и инсайтов, которые необходимо будет оценить.

Указанные методы продуцируют значительный объем информации: ценности и модели потребительского поведения; уровень и важность визуализации; отношение к изучаемой категории, уровень эмоциональной связи с категорией; типы дебютов в категории и степень и обусловленности.

Важно, что высказывания респондентов могут быть наложены на предварительно изученный материал этнографического этапа, что позволяет провести кросс-валидизацию и повысить качество выводов о том, какие концепции являются наиболее значимыми.

В процессе ведения тематических блогов, исследователь имеет возможность задавать узко направленные вопросы и просить производить визуализацию интересующих явлений, что позволяет впоследствии, оценить степень динамики изменений мнения потребителей (если таковая имеет место быть) и производить дифференциацию модельных точек зрения и артефактуальных.

Важно также отметить, что самостоятельный отбор респондентами этнографических материалов позволяет с высокой точностью определить приоритеты и социальные стереотипы потребительского поведения.

Этот этап исследования может реализоваться с помощью разнообразных блоговых ресурсов (ключевым, конечно, является Живой Журнал), социальных сетей, а также специализированного программного обеспечения. Исследователь сохраняет полный контроль над процессом, постоянно вмешивается в обсуждения респондентов и провоцирует их на дальнейшую рефлексию. Открытые платформы позволяют респондентам обращаться в своих рассуждениях к текстам других блоггеров, указывать, какие темы их больше всего интересуют, на какие объекты они обращают свое внимание. Все это дополняет картину жизненного мира и усиливает понимание исследователем мотивов выбора того или иного варианта концепции.

Особую роль играет такое сопровождение в ситуации, когда на тестирование выносятся объекты с высокой долей этической составляющей (рациональное потребление, «зеленое» потребление, социальные проекты). Тестирование подобных идей традиционно вызывает сложности, поскольку предполагает высказывания на социально-болезненные темы. Модератор стремится приложить все усилия для избегания мини-эффекта «спирали молчания»¹, что бывает нелегко в рамках открытой дискуссии.

Таким образом, использование методов self-ethnography позволяет в значительной мере преодолеть ограничения, накладываемые спецификой проведения качественного тестирования и повысить качество результатов. Компиляция методов коммуникативного и этнографического исследований усиливает эвристическую ценность исследовательской работы в рамках маркетингового проекта.

Потребительские стратегии российских граждан

Галицкий Ефим Борисович,
Петренко Елена Серафимовна,
Галицкая Елена Григорьевна
Фонд «Общественное Мнение»

Статья посвящена описанию типологии российских граждан с точки зрения из потребительских стратегий. Типология построена по данным массовых опросов, проведенных ФОМ по месту жительства респондентов.

Использовались, во-первых, объединенные данные шести еженедельных общероссийских репрезентативных опросов, проходивших в сентябре–октябре 2010 года. Суммарный размер выборки этих опросов составил 12 000 респондентов. Выборкой охвачено 205 населенных пунктов, расположенных в 64 субъектах Российской Федерации. Опросы репрезентируют взрослое (18 лет и старше) население страны в целом.

Во-вторых, использовались данные Мегаопроса, проходившего в рамках проекта ФОМ «Георейтинг» в ноябре 2010 года. Суммарный размер выборки каждой (в т.ч. ноябрьской) волны этого гигантского опроса составляет 34 000 респондентов. Выборкой охвачено 1954 населенных пункта, в том числе 676 городов, 315 поселков городского типа и 963 села. В отличие от еженедельных опросов, опросы проекта «Георейтинг» репрезентируют каждый из 68 субъектов РФ, в каждом из них опрашивается по 500 респондентов. На территории изучаемых 68 субъек-

¹ Ноэль-Нойман Э. Общественное мнение. Открытие спирали молчания: Пер. с нем. М.: Прогресс-Академия; Весь Мир, 1996. С. 16.

тов РФ проживает 91 % взрослого населения страны, так что и о стране в целом опросы «Гео-рейтинг» дают исчерпывающее представление. Более того, благодаря чрезвычайно большому суммарному размеру выборки, эти опросы позволяют изучать весьма свойства очень небольших групп российского населения. Например, даже группы, размер которых составляет 0,5% российского населения, представлены в выборке примерно 170 респондентами, то есть можно достоверно судить об их особенностях!

В ходе опросов, о которых шла речь выше, в анкеты был включен вопрос: «Скажите, что из перечисленного Вам доводилось делать за последние два-три года?» Для ответа респондентам предлагалось ознакомиться с карточкой, на которой перечислялись разного рода современные потребительские практики, в т. ч. брать кредит в банке; оформлять кредит в магазине; работать за компьютером; пользоваться интернетом, вести переписку по электронной почте; ездить за границу; расплачиваться за товары или услуги при помощи пластиковой карточки; иметь дело с иностранной валютой; пользоваться услугами косметических салонов; летать на самолетах; приобретать спортивные товары и / или туристическое снаряжение; заниматься в фитнес-центре, спортивном клубе; пользоваться услугой доставки товаров на дом; получать дополнительное образование, повышать квалификацию; пользоваться мобильным телефоном; пользоваться смартфоном.

Цель выполненного нами исследования состояла в том, чтобы изучить сегодняшнее состояние российского общества с точки зрения такого рода современных потребительских практик. Для достижения этой цели требовалось, во-первых, изучить структуру современных потребительских практик, выяснить, какие из них, как правило, сопутствуют друг другу. Зная такую структуру, можно описать потребительское поведение людей на особом обобщенном языке, адекватном современному этапу развития российского общества.

Во-вторых, надо было выявить и описать на указанном выше обобщенном языке характерные для современной России сочетания потребительских практик. Другими словами, надо было найти типичные стили современного потребительского поведения россиян.

Если говорить о методическом арсенале исследования, то это разработанная в ФОМ инновационная аналитическая стратегия анализа социальных структур, которую мы по аналогии с географией называем «социографией». Суть новой стратегии состоит в последовательном применении метода главных компонент (классического, или категориального), а также трех методов кластеризации: метода иерархической кластеризации (для отыскания оптимального числа кластеров и предварительного, «грубого» отыскания самих кластеров); метода К-средних для уточнения состава найденных кластеров и метода классификационных деревьев для получения краткого описания кластеров. Благодаря применению этого метода становится возможным говорить не о кластерах, представляющих собой, по сути, как правило, достаточно аморфные сгущения точек в пространстве изучаемых переменных, а типах респондентов. Каждый тип респондентов описывается набором простых логических правил. Тип представляет собой объединение подтипов, а каждый подтип объединяет в себе респондентов, которые в точности одинаково ответили на определенные, ключевые вопросы анкеты.

Итак, аналитическая стратегия ФОМ, представляет собой сочетание ряда методов, реализованных, в частности, в программном комплексе SPSS. Надо отметить также, что ряд методов используется не совсем обычным образом, стратегия предполагает использование определенных правил, своего рода *know how*. Например, кластерный анализ, как правило, выполняется не в пространстве исходных переменных, а в пространстве факторов, причем каждый фактор предварительно умножается на корень квадратный из соответствующего ему собственного числа. Это позволяет вернуть облаку точек, соответствующих отдельным респондентам и расположенных в факторном пространстве, ту форму, которую это облако имело в пространстве исходных переменных.

Как показало исследование, переменные, соответствующие отдельным потребительским практикам, естественным образом образуют четыре комплекса. Практики, входящие в каждый комплекс, тесно взаимосвязаны. Из содержательных соображений практики, входящие в первые два комплекса, можно условно считать практиками высокого уровня, а в два других комплекса — базовыми практиками.

Комплексы базовых практик имеют на сегодняшний день достаточно широкое распространение среди значительной части населения. К ним относятся *компьютерно-коммуникационные практики* (работа за компьютером, пользование интернетом и электронной почтой, пользование мобильным телефоном) и *кредитные практики* (получение кредитов в банке или в магазине).

Комплексы практик высокого уровня распространены в гораздо меньшей степени, для их освоения необходимо сравнительно высокое материальное положение. Первый из этих комплексов мы условно назвали комплексом практик, направленных на *развитие личности*. В него входят приобретение спортивных товаров или туристического снаряжения, занятия в фитнес-центре или спортивном клубе, пользование услугой доставки товаров на дом, пользование смартфоном, получение дополнительного образования или повышение квалификации. Второй комплекс практик высокого уровня составляют три сильно коррелирующие между собой практики: поездки за границу, пользование иностранной валютой и полеты на самолете. Этот комплекс практик мы будем условно именовать практиками, связанными с *поездками за границу*.

В ходе реализации процедуры факторного анализа для каждой из четырех групп переменных был построен фактор, обобщенно отражающий степень использования респондентами соответствующих потребительских практик. После умножения каждого фактора на корень из собственного числа в факторном пространстве было выявлено десять кластеров — десять относительно компактных сгущений точек, соответствующих респондентам. Благодаря применению метода классификационного дерева каждому кластеру удалось дать простое определение. Оказалось, что для представителей каждого кластера свойственен свой, вполне определенный стиль потребительского поведения.

Так, например, группа респондентов, получивших условное название «Стиль 1. Все современные практики» состоит из двух подгрупп. Первую подгруппу (1.1) составляют респонденты, которые упомянули не менее трех из пяти практик, направленных на развитие личности, не менее двух из трех практик, связанных с поездками за границу, и не менее двух из трех практик, связанных с компьютеризацией или коммуникациями. Размер этой подгруппы составляет 1,5 % населения страны. Вторую подгруппу (1.2) составляют респонденты, которые упомянули две из пяти практик, направленных на развитие личности, две из трех практик, связанных с поездками за границу, и не менее двух практик, связанных с компьютеризацией или коммуникациями. Размер этой подгруппы составляет 0,4 % населения страны. Таким образом, численность группы «Стиль 1» составляет 1,9 % от численности взрослого населения Российской Федерации, или 2,2 млн. чел.¹

Наличие точного определения групп и подгрупп в виде простых логических формул позволяет воспроизводить их по данным всех регулярных опросов ФОМ, а, следовательно, изучать динамику их численности и свойства.

Важно отметить, что потребительские практики высокого уровня характерны только для первых пяти групп респондентов (стили потребительского поведения с первого по пятый). Их суммарный размер — 20,4 % населения страны. Интересно, что, несмотря на наличие определенной случайной погрешности исследований, в точности эта величина была зафиксирована по данным двух независимо проведенных опросов: шести еженедельных опросов и Мегаопроса «Георейтинг». Остальные пять групп (стили 6–10) практиками высокого уровня почти совсем или совсем не пользуются. Относительные размеры групп населения с различными стилями потребительского поведения иллюстрирует рис. 1.

Аналогично выявляются особенности остальных десяти групп с различными стилями потребительского поведения. При этом различаются ключевые особенности и свойства, не являющиеся ключевыми. Под ключевыми особенностями выступают определяющие свойства, т. е. те сочетания значений признаков, по которым можно однозначно отличить представителей данной группы от всех остальных. Вторую группу составляют те свойства, которые у той или иной группы наблюдаются чаще, чем по населению в целом.

¹ Численность взрослого населения России по данным Росстата составляет 115,9 млн. чел.



Рис. 1. Структура стилей потребления населения России, %

Ключевые особенности групп выражаются, прежде всего, в терминах четырех обобщенных индексов потребительского поведения. Каждый индекс характеризует какой-либо один из четырех комплексов практик, которые были выявлены с помощью факторного анализа. Способ расчета индикатора, в отличие от фактора, очень прост. По каждому респонденту рассчитывается, сколькими практиками из данного комплекса он пользуется, и полученная сумма делится на число практик в данном комплексе. Например, если респондент отметил одну практику из данного комплекса, а всего к комплексу отнесено три практики, то индикатор будет равен одной трети. Таким образом, определение групп дается в терминах четырех индикаторов потребительского поведения: «Развитие личности», «Заграничные поездки», «Пользование кредитами» и «Компьютеризация и коммуникации». Наряду с этим, для «узнавания» представителей некоторых групп ключевую роль играют некоторые другие признаки, которые будут упоминаться по мере надобности.

Так, для первых пяти стилей характерны практики высокого уровня. Между этими стилями нет жесткой иерархии. Один стиль характеризуется одним набором таких практик, а другой — другим. Исключение составляет стиль 1: для его представителей характерен достаточно высокий (хотя, возможно, и не максимально возможный) уровень трех типов практик из четырех, то есть всех, кроме пользования кредитами. Думается, его представители зачастую не нуждаются в кредитовании в виду достаточно высокого материального положения.

Далее мы видим, что значения индикатора «Пользование кредитами» не играют ключевой роли для стиля 1. Эти практики могут, как присутствовать, так и не присутствовать у их представителей. Приведем диаграмму, иллюстрирующую средние значения четырех индикаторов для стиля 1 (рис. 2).

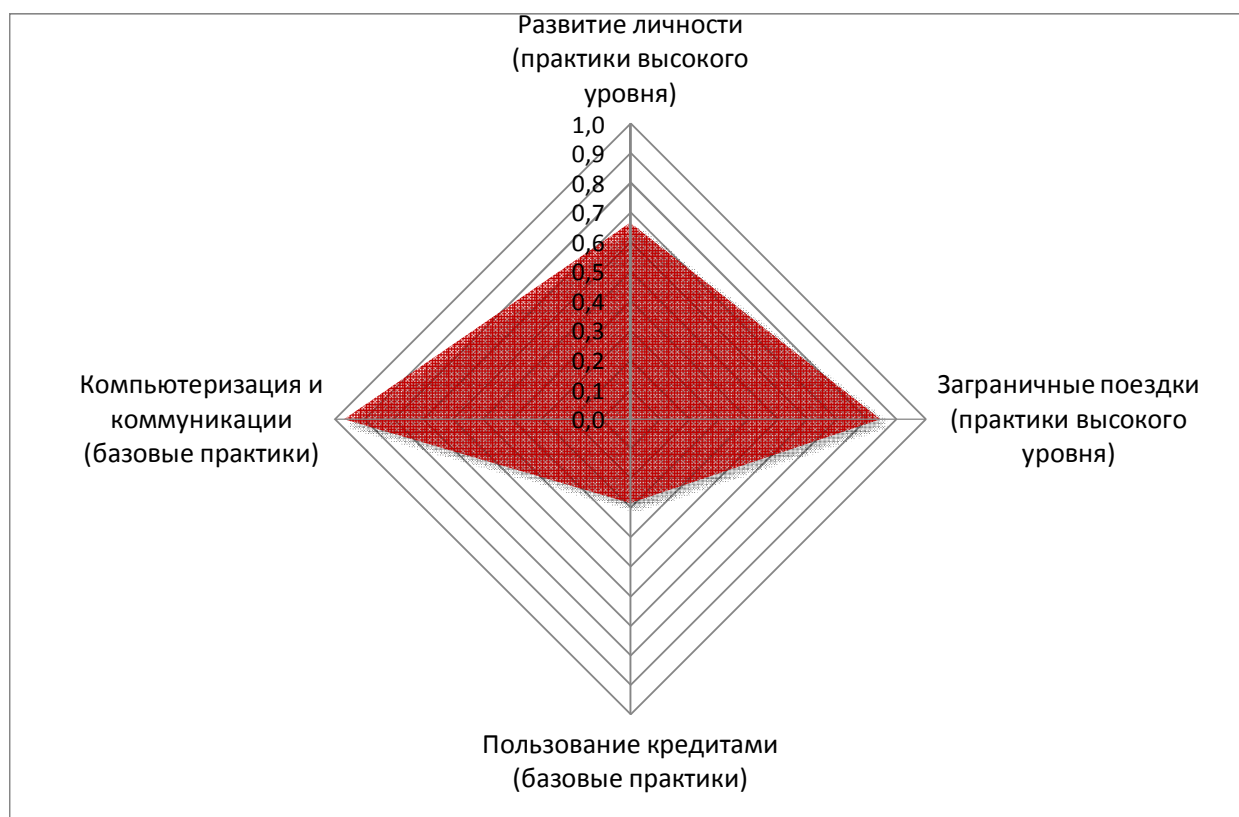


Рис. 2. Средние значения индикаторов потребительского поведения для группы «Стиль 1. Все современные практики».

Для отнесения респондента к группе «Стиль 1» достаточно, как уже отмечалось, пользования двумя практиками комплекса «Компьютеризация и коммуникации» из трех, однако подавляющее число представителей этой группы пользуются всеми тремя практиками. Среднее значение соответствующего индикатора почти равно единице (оно составляет 0,97), а среднее значение индикатора «Заграничные поездки» находится между 0,8 и 0,9 (оно равно 0,85) (см. рис. 2). Практики комплекса «Развитие личности» характеризуются значением индикатора 0,66, а комплекса «Пользование кредитами» — 0,29.

Рассмотрим особенности группы «Стиль 1» в плане социально-демографических, экономических характеристик и места жительства на фоне взрослого населения страны в целом (табл.).

Таблица

Доли лиц с разными характеристиками в составе группы «Стиль 1. Все современные практики» и среди взрослого населения России в целом, %

	Группа «Стиль 1»	Россия
Моложе 36 лет	71	36
Высшее образование	65	20
Доход более 20 тыс. руб.	39	8
Хватает на крупную бытовую технику	70	19
Руководители и специалисты	66	18
Москвичи	40	7

Мы видим, что лиц моложе 36 лет в этой группе почти вдвое больше, чем в населении в целом, лиц с высшим образованием — более чем втрое, лиц, чей доход превышает 20 тыс. рублей в месяц — почти впятеро и т.д. Двое из пяти представителей данной группы — москвичи, т.е. доля жителей столицы здесь почти в шесть раз выше их доли среди населения страны. Таким

образом, данный стиль поведения типичен, в основном, для успешных людей. Достаточно сказать, что подавляющее их большинство (70% против 19% по населению в целом) имеет материальную возможность для покупки, по крайней мере, бытовой техники¹. Аналогично все остальные потребительские стили имеют свои особые и четко проявленные социально-демографические профили.

В заключение приведем карту России, на которой разные субъекты федерации окрашены в зависимости от того, какую долю среди их жителей составляют представители групп с первыми пятью типами потребительского поведения, т.е. пользующиеся практиками высокого уровня (рис. 3). Мы видим, что субъектов федерации, в которых доля пользующихся практиками высокого уровня составляет треть и более, лишь несколько. Это Москва, Московская область, Санкт-Петербург, Ханты-Мансийский автономный округ, Еврейская автономная область, а также Амурская, Магаданская и Камчатская области. В противовес этому в Тверской, Тамбовской, Пензенской областях, а также в республике Мордовии первый–пятый потребительские стили свойственны менее чем для 9 % взрослых жителей. Таким образом, как в социальном, так и в территориальном отношении для нашей страны сегодня характерны очень сильные контрасты в плане потребительского поведения.

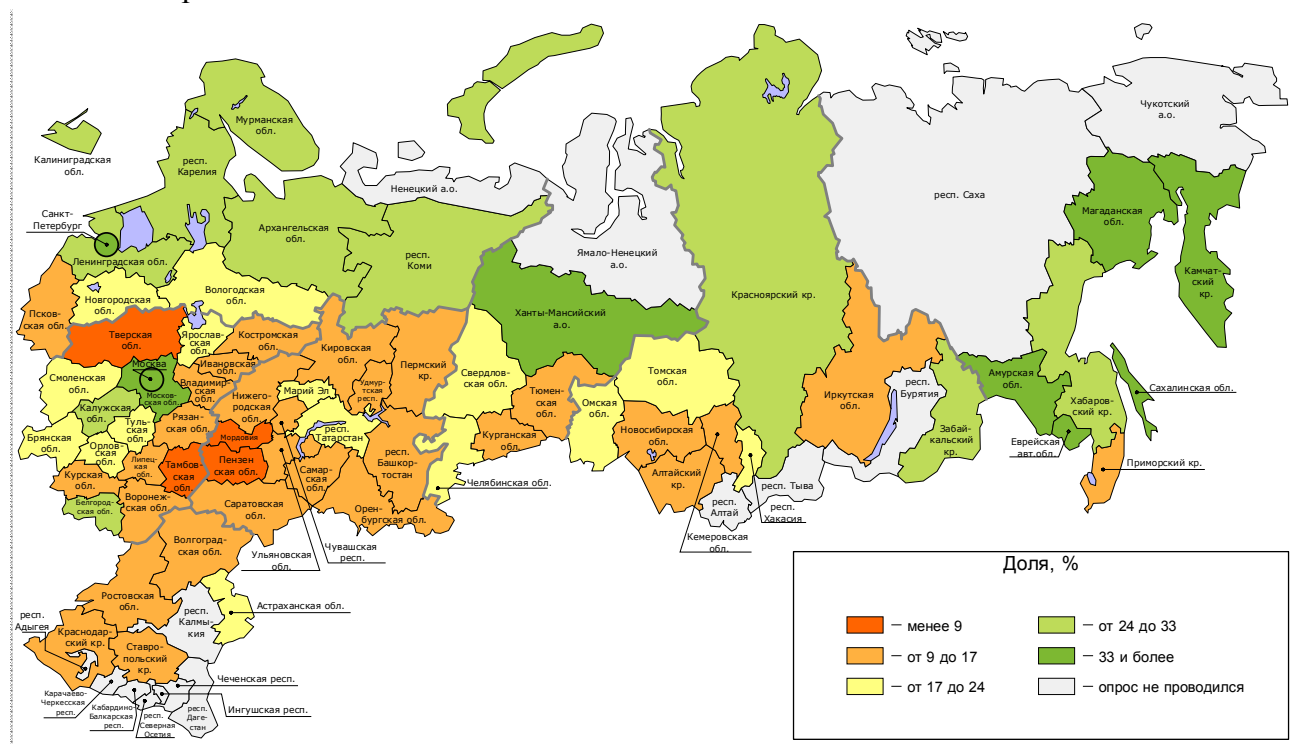


Рис. 3. Суммарные доли представителей групп с первыми пятью стилями потребительского поведения среди жителей субъектов РФ по данным опроса «Георейтинг», %

Подведем итоги. С помощью специальной методики, разработанной в Фонде Общественное Мнение, был проведен углубленный анализ данных опросов населения. В результате, во-первых, была выявлена структура современных потребительских практик. Как выяснилось, современные потребительские практики делятся на четыре группы: практики, направленные на развитие личности, практики, связанные с поездками за границу, компьютерно-коммуникационные практики и кредитные практики. Практики первого и второго типа получили название практик высокого уровня, а третьего и четвертого — базовыми практиками.

¹ Материальное положение респондентов оценивалось по ответам на вопрос «Какое высказывание точнее всего описывает материальное положение Вашей семьи?» Предлагалось выбрать один из следующих вариантов ответа: 1) денег не хватает даже на питание; 2) на питание денег хватает, но одежду, обувь купить не можем; 3) на одежду, обувь денег хватает, но крупную бытовую технику купить не можем; 4) на бытовую технику денег хватает, но автомобиль купить не можем; 5) на автомобиль денег хватает, но квартиру или дом купить не можем; 6. на квартиру или дом денег хватает.

Во-вторых, анализ показал, что сегодня в России существует десять различных стилей потребительского поведения: от группы «Стиль 1. Все современные практики» до группы «Стиль 10. Нет современных практик». Только представители первых пяти групп пользуются теми или иными практиками высокого уровня, остальные россияне ими практически совсем не пользуются. Суммарный размер первых пяти групп составляет 20,4 % российского населения, причем данный результат с точностью до десятых долей процента зафиксирован по данным двух независимых массовых опросов населения.

Сопоставление типологии по стилям потребительского поведения с другой — ресурсной — типологией Фонда Общественное Мнение показало, что активность пользования современными потребительскими практиками напрямую зависит от объема разного рода ресурсов, которыми располагает человек. Показано также, что структура населения с точки зрения стиля потребительского поведения существенно различается в разных субъектах Российской Федерации.

Особенности организации опроса участников и посетителей выставки

Герасимова Ксения Георгиевна,
Самарский государственный университет

Согласно данным Российского союза выставок и ярмарок в 2009 году его члены провели свыше 1000 профессиональных мероприятий, в которых приняло участие более 100 тысяч экспонентов, а объем задействованных выставочных площадей приблизился к отметке в полтора миллиона квадратных метров¹. Речь идет о выставках как рыночных мероприятиях, на которых экспоненты представляют на основе выставочных образцов производимые товары и услуги². Подробнее о специфике выставок можно, например, найти в работе³. Цели участия в выставках могут быть самыми разнообразными, начиная с тестирования концепции новых товаров/услуг, изучения ожиданий потребителей и до поддержания/улучшения имиджа компании. Технология эффективного участия в выставках достаточно проработанная тема, представленная широким спектром специализированной литературы⁴.

Не менее интересна выставка сама по себе как продукт деятельности выставочной компании. Как и любое другое коммерческое предприятие, выставочный организатор, нуждается в постоянном притоке маркетинговой информации.

Строго говоря, все исследовательские задачи, стоящие перед выставочной компанией можно разделить на несколько групп, в том числе:

1. Изучение состояния той или иной отрасли, и потребности ее агентов/участников в выставке, как инструменте продвижения
2. Анализ конкурентной среды: прямых конкурентных продуктов – выставочных проектов аналогичной тематики, и косвенных – тех каналов продвижения, которые используются участниками того или иного отраслевого рынка.
3. Оценка успешности организации и других показателей выставочных проектов самой компании и другие.

Обратим особое внимание на последнюю задачу. Успешность организации выставочного проекта — комплексное понятие, и, безусловно, для каждой выставочной компании существует своя специфика ее оценки и учета. Вместе с тем, можно принципиально разделить на две большие группы, показатели, которыми она измеряется. Прежде всего, это строго объективные количественные данные, так называемая статистика выставки, которая включает в себя такие по-

¹ Алексеев С.П. Выставочная индустрия 2009. Как мы пережили кризис? // Экспо Ведомости. 2010. №2.

² Концепция развития выставочно-ярмарочной деятельности в Российской Федерации (Одобрена на заседании Правительства Российской Федерации, протокол № 12 от 29 марта 2001 г.) // URL: [http://www.vlc.ru/forum/con_forum/0290301.htm]

³ Лебедева К.Г. К определению понятия выставки // Материалы XI международной конференции «Культура, личность, общество в современном мире: методология, опыт эмпирического исследования». Екатеринбург: Миниполаборатория факультета политологии и социологии УрГУ, 2008. Ч. 2. С. 174–176.

⁴ Добробабенко Е.В., Добробабенко Н.С. Выставка «под ключ». Готовые маркетинговые решения. СПб.: Питер, 2007; Александрова Н.В., Сорокина Е.В., Филоненко И.К. Участие в выставке: формула успеха. М.: ПроНекст, 2006; Беляновский А.С. Экспонент, помоги себе сам! М.: ТПП-Информ, 2010 и др.

казатели как количество посетителей и участников, задействованная выставочная площадь, финансовые показатели, число стран участниц и пр. В отношении данной группы показателей традиционно используется система выставочного аудита, признанная в профессиональном сообществе¹.

Формируя общую картинку в абсолютных показателях, данные выставочного аудита составляют только часть общей картины. Не менее важно получить и так называемые «качественные» показатели — субъективные оценки посетителей и участников выставки. Наиболее адекватный метод получения такого рода информации выступает опрос.

Рассмотрим наиболее важные, с нашей точки зрения, аспекты, определяющие специфику проведения исследовательских работ для решения маркетинговых задач организаторов выставки. Помним, что выставка, как продукт имеет две принципиально различные группы потребителей: участники (экспоненты) и посетители. Именно эта двойственность будет определять двусторонний подход к некоторым вопросам, возникающим в ходе проведения исследования.

Специфические задачи исследования

Итак, как уже отмечалось в ходе опроса участников и посетителей на выставке мы имеем возможность получить информацию. Исходя из подхода к пониманию успешности выставки и критериев, которые мы в него закладываем, а также наших потребностей, в частности, проектной группы, которая отвечает за организацию каждой конкретной выставки, формируются показатели и соответствующие им индикаторы. Наметим наиболее универсальные показатели.

1. Показатели, относящиеся к посетителям выставки:

- социально-демографический и поведенческий профиль (в зависимости от тематики выставки признаки данной группы могут варьироваться);
- цели посещения выставки;
- источники получения информации о выставке;
- удовлетворенность выставкой в целом и различными ее аспектами (спектром участников, развлекательной программой, выставкой в целом и т.п.);
- предложения по модификации выставочного продукта и т.д.

В зависимости от задач исследования также могут быть включены самые разнообразные блоки, связанные с тестированием нового названия выставки, восприятию цен на входные билеты, наиболее запомнившуюся часть развлекательной программы и т.д. Целесообразно по возможности максимально формализовать вопросы анкеты, отдать предпочтение закрытым вопросам. С одной стороны, это значительно облегчит обработку и анализ результатов, с другой позволит получить интересующую нас информацию в сжатые сроки. При разработке инструментария также необходимо учитывать показатели, относящиеся к участникам выставки:

- профиль участников выставки;
- цели участия в выставке и место выставок в общей стратегии продвижения компании;
- источники получения информации о выставке;
- оправданность ожиданий и удовлетворенность различными аспектами выставочного продукта (количеством и качеством посетителей, работой персонала, дополнительными мероприятиями и т.д.);
- предложения по модификации продукта;
- намерение принимать участие в выставке следующего года и др.

Задачи опроса участников выставки, также могут корректировать и дополнять спектр измеряемых признаков.

Выборка

Репрезентативность выборочной совокупности — принципиальный методологический вопрос при организации классического социологического и маркетингового исследования². В на-

¹ Тищенко В. Аудит выставок, шаг 2: качественный состав посетителей // Экспо Ведомости. 2007. № 1.

² См. например в работах: Голубков Е.П. Маркетинговые исследования // Маркетинг в России и за рубежом. 2001. № 1; Давыдов А.А. Репрезентативность выборки // Социологические исследования. 1990. № 1.

шем же случае, специфика ситуации диктует и дополнительные моменты, которые необходимо учитывать.

Если в ситуации опроса участников (экспонентов) мы можем не рассматривать вопрос с выборкой: поскольку интервьюированию могут подвергнуться все участники выставки, присутствующие в выставочном зале, то есть очные участники. Здесь однако необходимо помнить, кто именно из представителей компании станет нашим респондентом — простой «стендист» (сотрудник компании, работающий на выставке, как правило, менеджер по продажам), маркетолог или, в конечном счете, тот, кто от лица компании уполномочен принимать решения об участии в выставках. При анализе результатов опроса участников, на наш взгляд, принципиально учитывать статус представителя компании, принимающего участие в опросе.

Не стоит забывать также о, так называемых, заочных участниках выставки — компаниях, которые не представлены самостоятельными стендами или представителями, а как бы косвенно (через рекламно-информационные материалы) участвуют в выставке. Относительно данной группы опрос проводится по истечении некоторого времени по завершении выставки, поскольку данные компании не могут оценить свое участие, что называется «на месте».

С посетителями выставки вопрос относительно выборочной совокупности встает наиболее остро. Прежде всего, в отношении того, что до завершения выставки мы окончательно не знаем объемы генеральной совокупности — общего количества посетителей нашей выставки. Можно сказать, что эта аудитория формируется параллельно с проведением выставки, а значит и с проведением самого исследования/опроса. Причем, выставки, как правило, проходят в течение нескольких дней, и количество посетителей в течение каждого дня может значительно различаться. Например, для потребительских выставок приток посетителей усиливается в выходные дни, в то время как для специализированных (проводимых, как правило, в будни) наибольшее число посещений приходится на первые дни выставки.

Какие же могут быть пути решения этой проблемы? Для выставочных проектов, которые успешно проводятся уже не первый год, вполне информативными будут данные о посещаемости в предыдущие годы, с учетом их динамики, а также корректировки стратегии продвижения самого выставочного проекта применительно к данному году.

Для расчета объема выборочной совокупности в социологических исследованиях применяются соответствующие формулы¹, основанные на данных о генеральной совокупности. Как мы уже отмечали, априорное знание о размере генеральной совокупности посетителей выставки нам не дано, поэтому здесь мы можем ориентироваться на правила, предлагаемые FKM (Обществом Независимого Контроля Выставочной Статистики, Германия), которые рекомендуют опрашивать не менее 300 респондентов на выставочном мероприятии с общим числом посетителей до 5000 человек, от 5000 до 25 000 — не менее 500 человек, от 25 000 до 50 000 — не менее 2 %, свыше 50 000 — не менее 1000 чел.²

Тип формирования выборки возможен лишь случайным образом: интервьюер отбирает респондентов так, чтобы каждый из посетителей выставки имеет равный шанс попасть в выборочную совокупность.

Ситуация опроса

Выставка для организатора является продуктом, товаром и, в то же время, исследовательским полем, в рамках которого можно оценить, насколько этот продукт востребован и отвечает запросам ключевых потребителей: участников и посетителей выставки.

При опросе посетителей выставки интервьюер должен строго придерживаться случайности отбора респондентов. Однако задачи исследования предполагают, что у респондентов уже сформировалось мнение о посещаемой ими выставке, поэтому проводить опрос только что вошедших в выставочный зал посетителей нецелесообразно, мы просто не получим ответы на наиболее важные для нас вопросы. Таким образом, нам нужны те, кто уже ознакомился с боль-

¹ Паниотто В.И. Максименко В.С. Количественные методы в социологических исследованиях. Киев, 2003. С. 167–180.

² Александрова Н.В., Филоненко И.К. Выставочный менеджмент: стратегии управления и маркетинговые коммуникации. Тула: ОАО «Лев Толстой», 2006. С. 310.

шей частью экспозиции, достаточно длительное время пробыл на выставке и сформировал для себя относительно устойчивое представление о ней. Безусловно, выставки различаются по своему масштабу, во многом ситуацию опроса формирует физическое пространство выставочного центра. На наш взгляд, наиболее адекватным решением будет опрашивать респондентов на выходе с экспозиции.

Опрос участников выставки более комфортен, поскольку респонденты «физически закреплены» за своими стендами и в данном случае мы уже можем позволить себе полуформализованное интервью.

Для формирования мнения относительно успешности своего участия в выставке компаниям недостаточно нескольких часов, поэтому опрос участников целесообразно начинать на второй-третий день выставки.

Оговоримся также, что выставки обладает, так называемым отложенным эффектом: результаты от участия в выставке могут проявляться (и проявляются) уже после ее завершения и оценить эффективность, успешность участия становится возможным гораздо позже. Например, переговоры, проведенные на выставке с потенциальными партнерами, приводят к заключению долгосрочных контрактов через несколько месяцев, или, посетив автомобильную выставку и сравнив несколько автомобильных марок в одном месте, потребитель через некоторое время приходит в автосалон и приобретает машину. В связи с этим относительно участников необходимо отслеживать ситуацию после выставки. Наиболее целесообразное решение — проведение телефонного опроса через 3–5 месяцев (в зависимости от отрасли) по прошествии выставки.

Глубина анализа данных

Полученная в ходе опроса посетителей и участников информация позволяет увидеть выставочный продукт «глазами потребителей»: выявить их мнение относительно участия/посещения выставки, сформировать представление о профиле посетителей и т.д. Как правило, для принятия управленческих решений на основе результатов опроса посетителей и участников выставки организаторы используют «простые» данные описательных статистик — частотные распределения, таблицы сопряженности и меры центральной тенденции. На «математические изыскания» просто нет времени и по большому счету необходимости. Данные, полученные по единым шкалам в течение нескольких лет реализации проекта, позволяют увидеть картину в динамике.

Помимо решения внутренних задач выставочной компании, например, по модификации продукта (выставки), результаты опроса участников и посетителей могут стать «элементом продвижения» самой выставки. Так, принимая решение об участии в выставке, компании могут обратиться к ним и узнать мнение экспонентов (насколько они удовлетворены участием, достигли ли своих целей и т.д.), оценить профиль посетителей (в какой степени он соответствует целевой аудитории компании) и т.д.

Проведение опроса участников и посетителей выставки, безусловно, имеет свою специфику: условия формирования выборки, обусловленность временных рамок исследования временем проведения выставки и др. Однако данные, полученные в ходе исследования, обладают колоссальной ценностью для выставочной компании: они позволяют оценить качество и при необходимости скорректировать продукт (выставки), формировать прогнозы, выстраивать контакт с целевой аудиторией, что является залогом успешной деятельности современной, ориентированной на маркетинг выставочной компании.

Результаты мониторинга социальных медиа в маркетинговых исследованиях

Давыдов Сергей Геннадьевич,
Стахмич Анна Сергеевна
*Международный институт маркетинговых
и социальных исследований «ГфК-Русь»*

Под социальными медиа обычно понимаются формы коммуникации, основанные на контенте, созданием которого занимаются сами пользователи. Как правило, их противопоставляют индустриальным, или профессиональным СМК. Безусловно, коммуникативные ситуации, при

которых медийный контент носит интерактивный характер и создается не специалистами, сами по себе далеко не новы. В то же время значительным импульсом к распространению непрофессиональных коммуникативных форм стало развитие онлайн коммуникаций. Поэтому социальные медиа в большинстве случаев относят к новым средствам массовой коммуникации, в свою очередь противопоставляемым традиционным масс-медиа, таким, как печатная периодика, радио или телевидение.

Едва ли кто-либо из специалистов в области маркетинга станет сегодня оспаривать тот факт, что социальные медиа на многих рынках являются значимой площадкой формирования потребительского поведения. Действительно, результаты многочисленных исследований свидетельствуют о том, что сбор информации в Интернете — важнейший вид деятельности, предшествующий покупке, на многих потребительских рынках. При этом наряду с официальной информацией от производителей и торговых ресурсов, покупателей интересует мнение обычных пользователей, имеющих опыт потребления данного товара или услуги и способных выступить в роли советчика. Размещение в Интернете сенсационных, скандальных материалов может оказать на коммерческую судьбу продукта существенное как положительное, так и отрицательное влияние.

В связи с появлением новой коммуникативной среды, использующейся, в том числе, для обсуждения товаров и услуг, появляются и маркетинговые инструменты, позволяющие работать с ней. Один из таких инструментов — средства мониторинга упоминаний, предназначенные для поиска, отслеживания и анализа определенного рода сообщений в социальных медиа. Как отмечает А. Рокина¹, для работы с англоязычным контентом подобных систем существует более 150; для мониторинга русскоязычных ресурсов — 8. В частности, это такие системы, как Brandspotter, Buzzware, IQBuzz, Kribrum, Monitorix, Semanticforce, Wobot и Youscan. Упомянем также бесплатные сервисы «Yandex.Блоги» и «Google.Блоги», которые позволяют осуществлять отбор упоминаний по ключевым словам, однако не снабжены каким-либо серьезным аналитическим функционалом.

Подобно традиционным поисковым онлайн системам, таким как Яндекс или Google, системы мониторинга социальных медиа также осуществляют отбор текстов по заданному запросу. Вместе с тем между данными системами существуют определенные различия. Например, если в первом случае запрос может быть сформулирован на естественном языке, то в последнем необходимо использовать логические операторы.

Другое отличие состоит в том, что поисковые системы обычно ориентированы на обработку запросов в режиме реального времени. В этом случае набор получаемых нами ссылок — своего рода «моментальная фотография» состояния онлайн ресурсов, содержащих определенные признаки. Безусловно, подобная задача, т.е. анализ текущего среза упоминаний, может ставиться и при изучении социальных медиа. Однако в маркетинговой практике чаще бывает востребован не моментальный информационный срез, а именно мониторинг ситуации, позволяющий отслеживать появление упоминаний если не в реальном времени, то с задержкой, сведенной к минимуму. Во-первых, эта возможность полезна для оперативного реагирования на появляющиеся упоминания. Примеров полезности данного инструмента можно привести немало — от своевременного реагирования на негативные сообщения, несущие репутационную угрозу, до отслеживания проблемных ситуаций в качестве стратегии поиска клиентов. Во-вторых, социальные платформы — «живые» и весьма динамичные системы, информация в которых не только постоянно появляется, но также и исчезает. Пользователи удаляют свои записи или прячут «под замок» (т.е. ограничивают к ним доступ, что делает публикации недоступными для мониторинговых систем), приостанавливаются действия аккаунтов и т.д. Об актуальности данной проблемы свидетельствует пример исследования LADA Kalina, который разбирается ниже. Мониторинг по данному запросу осуществлялся в течение месяца. Еще через месяц была проведена проверка доступности всех найденных ссылок. В итоге доступными оказались 86 % от исходного числа публикаций, причем на разных ресурсах этот параметр колебался от 57 до 100 % (см.

¹ Рокина А. Автоматические системы мониторинга соцмедиа: Презентация на конференции i-COMference-2011. Москва, 10 февраля 2011 г. // URL: [<http://rocid.ru/files/events/icom11/presentations/10feb.4--rokina.zip>]

Рис. 1). Отсюда можно сделать вывод, что, если мониторинговая система не позволяет сохранить полнотекстовые версии упоминаний в процессе отбора, обработку ссылок следует производить по мере их поступления во избежание потерь материала.

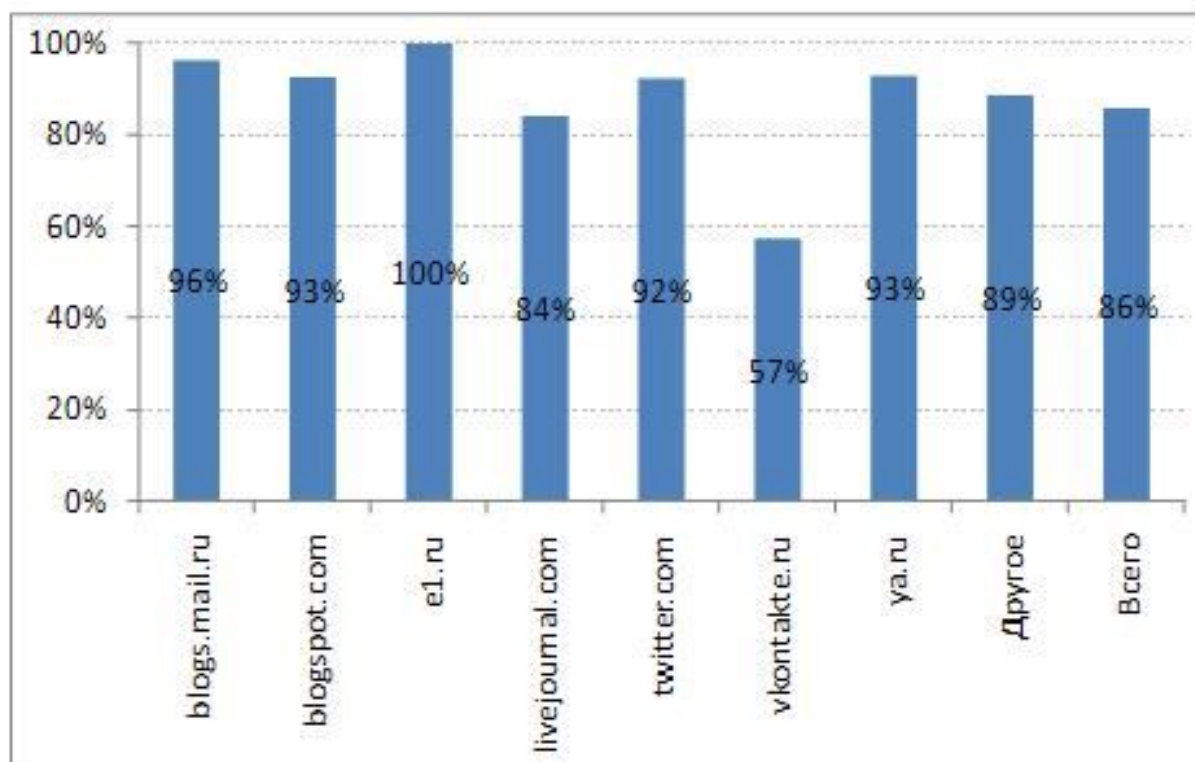


Рис. 1. Доступность ссылок в исследовании LADA Kalina через месяц после проведения мониторинга.

Ключевым вопросом, связанным с интерпретацией результатов мониторинга социальных медиа, является их репрезентативность. Репрезентируют ли данные упоминания какую-либо социальную реальность в той мере, в какой, например, результаты опросов общественного мнения или фокус-групп репрезентируют мнения или оценки тех или иных групп населения? Конечно, социальные медиа могут использоваться в качестве площадок для изучения общественного мнения и потребительского поведения. В таких случаях подбор респондентов и процесс сбора информации осуществляется под контролем специалистов и в соответствии с правилами проведения маркетинговых исследований, что позволяет обеспечить репрезентативность данных. Что же до результатов мониторинга публикаций в социальных сетях, то их, на наш взгляд, следует рассматривать исключительно в контексте саморепрезентации. Упоминания в социальных медиа — достаточно сложный информационный конструкт, формируемый профессионалами и непрофессионалами, лицами вовлеченными и невовлеченными (заинтересованными и незаинтересованными), преследующими различные цели и использующими различные коммуникативные технологии. Действительно, определенные сообщения, мнения, точки зрения, присутствующие в социальных медиа, совсем не обязательно будут также представлены на значимом уровне в массовом сознании или индустриальных медиа, и наоборот.

Определенным ограничением метода является то, что выявленные в ходе мониторинга упоминания рассматриваются в качестве анонимных и безадресных. Это вовсе не означает, что все интеракции в социальных медиа анонимные и безадресные; более того, знание автора и адресата реплики может существенным образом сказываться на восприятии сообщения. Однако аккаунты, принадлежность которых определенной публичной персоне широко известна и докумен-

тально подтверждена, составляют незначительную долю от общего числа регистраций. Более широкое установление взаимосвязи аккаунтов с их владельцами фактически является самостоятельной исследовательской задачей, которая к тому же не всегда может быть доведена до конца. Впрочем, это и не требуется, так как модель поиска референций по ключевым словам в большинстве случаев предполагает, что читатель, не вовлеченный в диалог изначально, будет воспринимать сообщение без привязки к автору и адресату.

Из сказанного выше непосредственно вытекает, что непрофессиональный характер публикаций в социальных сетях — не более чем допущение, часто не имеющее под собой никаких оснований. Более того, могут быть приведены многочисленные примеры именно профессиональной активности в социальных сетях — как открытой, так и скрытой. В некоторых случаях профессиональные публикации снабжены необходимыми атрибутами, позволяющими их идентифицировать, однако так бывает далеко не всегда.

Еще одно ограничение состоит в том, что в рамках мониторинга отбираются только сообщения, в которых непосредственно фигурируют заданные для поиска ключевые слова. Таким образом, ветви дискуссий, имеющие непосредственное отношение к предмету изучения, могут оказаться исключены из анализа.



Рис. 2. Контексты упоминания LADA Kalina.

Проиллюстрируем возможности маркетингового анализа текстов в социальных медиа опытом изучения упоминаний бренда LADA Kalina на русскоязычных дискуссионных онлайн-площадках. Исследование было проведено в декабре 2010 – январе 2011 г. отделом медиаисследований ГфК-Русь на базе системы YouScan — поисковой машины, которая агрегирует найденные упоминания в единую ленту, позволяет кодировать тональность, оставлять заметки, экспортировать результаты в единую таблицу в Excel и др. Выявленные при помощи YouScan тексты подверглись процедуре ручного кодирования по тематической принадлежности и тональности упоминания. Остановимся на некоторых полученных результатах.

Мониторинг высказываний, посвященных LADA Kalina, осуществлялся с 1 по 31 декабря 2010 г. За это время было зафиксировано 3069 содержательных упоминаний рассматриваемой

марки автомобиля в разнообразных контекстах, среди которых наиболее часто фигурируют четыре контекста; их описание приводится ниже.

1. Политический контекст (17 % упоминаний), попадание в который данной марки автомобиля в значительной степени связано с усилиями руководства страны по ее продвижению. В частности, напомним, что в конце августа 2010 г. В. Путин совершил на LADA Kalina поездку по новой трассе Чита – Хабаровск. Естественно, обсуждение этой поездки в анализируемый период успело сойти на нет. Однако в определенном смысле результатом поездки стало использование марки машины в контексте обсуждения актуальных проблем политического и социального характера: проблемы российской промышленности, взяточничество и т.д. Приведем пример высказывания из данной категории:

Министерство обороны разместило госзаказ на аренду и обслуживание 553 автомобилей на общую сумму 10 млрд. руб. ... А почему их не устраивают желтые LADA Kalina?

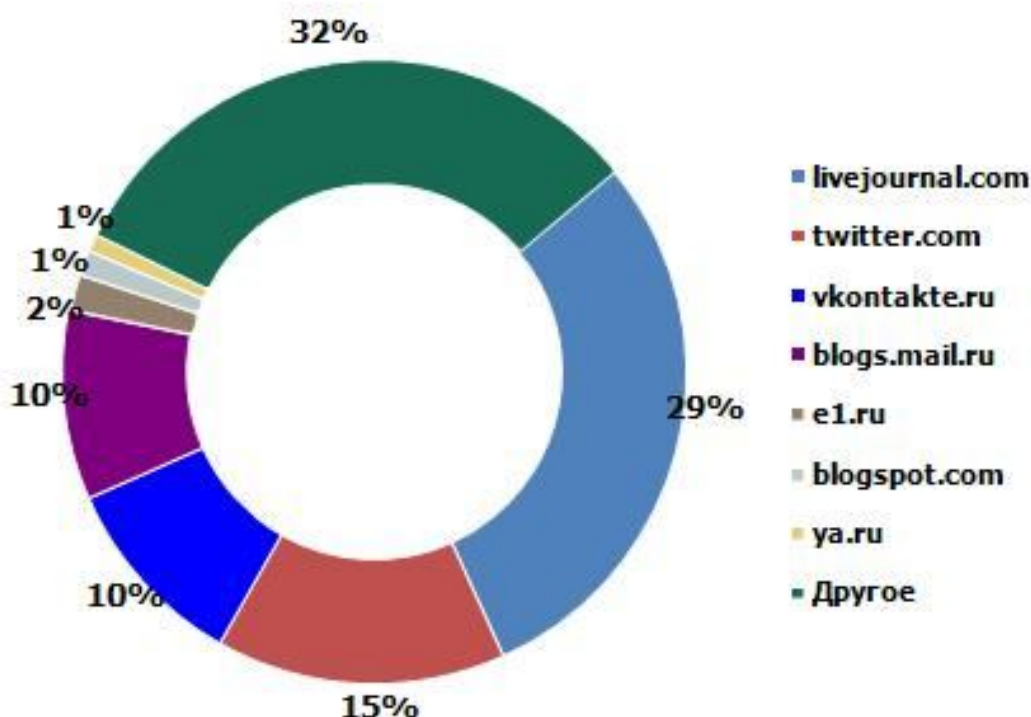


Рис. 3. Информационные площадки упоминания LADA Kalina.

2. Индустриальный контекст (23 % упоминаний) — обсуждение марки в связи с новостями российского автопрома. Пожалуй, одной из наиболее цитируемых и тиражируемых в социальных медиа новостей в данной категории оказалась информация от 14 декабря о выпуске серии автомобилей LADA Kalina Sport цвета «премьер» (то есть, желтого), точно таких, как была у В. Путина во время памятной поездки.

3. Фольклорный контекст (28 % упоминаний) — многочисленный шутки, анекдоты, стихотворения. Приведем наиболее популярную шутку:

Если верить показаниям спидометра, LADA Kalina стоит в гараже со скоростью 20 км/час.

4. Потребительский контекст (18 %) — отзывы об опыте эксплуатации и поломках LADA Kalina, информация об авариях, произошедших с ее участием, и т.д. Высказывания такого рода можно найти по отношению к любой более-менее массовой автомобильной марке. Опять же, приведем пример упоминания такого рода:

Ехал домой на Ладе Калине, не знаю, за что хают эту машину... вместительный салон, так все аккуратно, мягко бежит, хорошая машина, в общем.

Еще 14 % упоминаний вошли в категорию «Другое». Это публикации, либо относящиеся одновременно к нескольким контекстам (например, образцы политического юмора), либо не сводимые ни к одному из них.

Наиболее активно LADA Kalina обсуждалась на таких площадках, как LiveJournal (29% упоминаний), Twitter (15%), VKontakte.ru (10%) и Blogs.mail.ru (10%). На прочие ресурсы, коих было выявлено 639, приходится 36% упоминаний. При этом различные информационные поводы, связанные с рассматриваемой маркой автомобиля, неравномерно обсуждались на тех или иных сайтах. Например, статья в газете «Ведомости» от 2 декабря 2010 г. «LADA на троих»¹, в которой, помимо прочего, были озвучены планы о начале выпуска на базе LADA Kalina с 2013 г. бюджетных моделей Renault и Nissan, больше всего обсуждалась на ресурсах, не вошедших в первую четверку по числу упоминаний. Сатирическое стихотворение Д. Быкова «Азбучное», опубликованное в «Новой газете» 24 декабря², перепечатывалось, цитировалось и обсуждалось преимущественно пользователями LiveJournal; в нем также фигурирует анализируемая марка. (*«Ну вот, подошли к середине стишков, вторая пошла половина: на «Л» — утерявший доверье Лужков и желтая «Лада Калина».*) Однако наиболее успешным в плане цитируемости оказался упомянутый нами выше пресс-релиз о производстве LADA Kalina цвета «премьер». Данное сообщение активно обсуждалось на всех площадках и во всех перечисленных контекстах. В результате 14 декабря 2010 г. было зафиксировано 210 релевантных упоминаний LADA Kalina — рекордный показатель за период анализа.

Подведем общие итоги по результатам нашего исследования. Неоднозначное отношение к LADA Kalina русскоязычных пользователей социальных медиа в значительной степени связано с социальными аспектами потребления автомобильной техники в России, позицией на рынке данной марки и отечественного автопрома в целом. Наличие в собственности автомобиля по-прежнему остается в обществе признаком определенного уровня достатка и социального статуса. LADA Kalina, будучи для россиян одной из самых доступных автомобильных марок, таким образом, «по умолчанию» является объектом стремления для определенных групп населения. В этом смысле показательно, что обладатели LADA Kalina в большинстве случаев позитивно оценивают свой опыт вождения этого автомобиля.

На восприятие LADA Kalina накладывается негативный имидж отечественного автопрома, «визитной карточкой» которого, благодаря значительному информационному присутствию, она на сегодняшний день является. Большинство отрицательных упоминаний основаны не на фактах, а на предположениях, в основе которых лежит общее убеждение, что переход с российского автомобиля на иномарку также является шагом вверх по социальной лестнице. Опять же показательно, что LADA Kalina, как наименее престижная марка, часто противопоставляется автомобилям Mercedes и BMW, выступающим в качестве наиболее престижных среди массовых автомобилей. Также Lada Kalina — не просто самая популярная отечественная марка, на которую проецируется отношение к отечественному автопрому, но одна из «икон» продукции, сделанной в России. Она уступает зарубежным аналогам, порой заслуживает резкие негативные оценки, однако, что характерно, воспринимается при этом в качестве «своей». Это доказывают многочисленные образцы юмора, городского фольклора, а также иронические высказывания, встречающиеся в Интернете.

Определение функции распределения срока службы компьютера при помощи онлайн-опросов

Делицын Леонид Леонидович
ЗАО ФИНАМ

Количество персональных компьютеров (ПК) на душу населения страны — один из ключевых показателей развития информационного общества, используемый при составлении страновых рейтингов ООН и Международного союза электросвязи (МСЭ). В нашей работе представ-

¹ Фиалко А., Лысова Т. Lada на троих // Ведомости. 2010. № 228 (2746).

² Быков Д. Азбучное // Новая газета. 2010. № 145.

лены оценки размеров российского компьютерного *парка* и компьютерного рынка России в натуральных величинах. Чтобы определить неизвестную нам функцию распределения срока службы компьютера, ежегодно с 2004 по 2008 года проводился масштабный *онлайн-опрос* о сроке службы компьютера на Интернет-портале Rambler. Предварительные итоги первого опроса были доложены автором осенью 2004 г. на семинаре А.О. Крыштановского. Итоги моделирования обсуждаются данной работе, сопоставляются с новыми данными о продажах компьютеров и аналогичными прогнозами Министерства экономического развития (МЭР) и Министерства связи и коммуникаций.

Публикуемые различными статистическими источниками данные о продажах и парке компьютеров в России не согласуются между собой. Проиллюстрируем это утверждение при помощи простых расчетов. Запишем уравнение баланса парка компьютеров¹

$$Q(t) = N(t) + R(t). \quad (1)$$

Здесь $Q(t)$ — продажи ПК (штук), накопленные за весь период существования рынка (в РФ мы условно начинаем отсчет с 1986 года); $N(t)$ — парк ПК, который мы определяем как количество компьютеров, которые используются владельцами; $R(t)$ — накопленные продажи ПК на замену выбывшим из строя, они равны количеству вышедших из строя ПК.

На рис. 1(а) изображены накопленные продажи $Q(t)$, полученные суммированием продаж за каждый год (рис. 1d), а на рис. 1(b) показан парк ПК $N(t)$ за период с 1995 до 2010 г. и прогнозные значения МЭР на 2011 и 2012 годы. При анализе объема продаж ПК в России в натуральном выражении мы используем осреднение данных трех источников, в числе которых исследовательские компании IDC Russia, IT-Research и Gartner-Dataquest, публикующие статистику продаж новых компьютеров². Добавлена наша оценка продаж за 2010 год и наш наиболее оптимистический прогноз на 2010–2012 гг. При построении Рис. 1(а) и (d) использованы данные IDC, IT-Research, Gartner Dataquest, рис.1 (b) — данные Минкомсвязи РФ³ и прогноз МЭР⁴. На рис. 1(с) все данные использованы совместно. Открытыми символами показаны прогнозные значения. Сплошной линией на рис. 1(а) и (d) показаны экспоненты, наилучшим образом приближающие продажи за 1999–2007 гг. и отражающие известный факт роста российского рынка ПК устойчивыми темпами прироста около 21% в эти годы.

Переписав уравнение баланса парка компьютеров в виде

$$R(t) = Q(t) - N(t),$$

мы получаем формулу для вычисления парка вышедших из строя компьютеров. Вычисления приводят нас к неожиданному результату, который показан на рис. 1 (с). А именно, в 2009 году *накопленное* количество сломанных компьютеров $R(t)$ якобы уменьшилось на 6,8 миллионов штук. Однако величина $R(t)$ не может убывать, поскольку является суммой последовательности положительных чисел, представляющих количество вышедших из строя компьютеров за последовательные годы (или иные интервалы времени). Даже если бы в 2009 году не сломался ни один компьютер, величина $R(t)$ оставалась бы постоянной, но не могла бы снизиться. Единственным объяснением снижения накопленного числа сломанных компьютеров был бы массовый ремонт вышедших из употребления устройств. Однако для объяснения столь крупных расхождений потребовался бы внезапный ремонт 6,8 миллионов ПК в 2009 году, а столь масштабные работы вряд ли остались бы незамеченными. Другой причиной расхождений могли бы быть большие отрицательные ошибки в оценке рынка ПК аналитическими агентствами, однако

¹ Васильев В.В., Салютин Т.Ю. Мониторинг информатизации: показатели, методология оценки и прогнозирования. М.: Палеотип, 2005.

² IT Research: российский рынок десктопов «просел» на 30% // Режим доступа: <http://www.itoday.ru/news/11109.html>; IT-research подводит итоги 2009 г. и дает прогноз на 2010 г. // Режим доступа: <http://www.itresearch.ru/press-center/news/detail.php?ID=13310>

³ Кузовкова Т.А., Тимошенко Л.С. Анализ и прогнозирование развития инфокоммуникаций. М.: Горячая линия – Телеком, 2009.

⁴ Министерство экономического развития. Прогноз социально-экономического развития Российской Федерации на 2010 год и на плановый период 2011 и 2012 годов // Режим доступа: <http://www.economy.gov.ru/minec/activity/sections/macro/prognnoz/doc1254407742765>

падение рынка ПК, начиная с последнего квартала 2008 года, отмечалось не только в России. Более вероятным нам кажется предположение о том, что претерпела изменения методика Минкомсвязи оценки величины парка ПК, причем смыкание данных было произведено без учета совместимости с данными о продажах ПК.

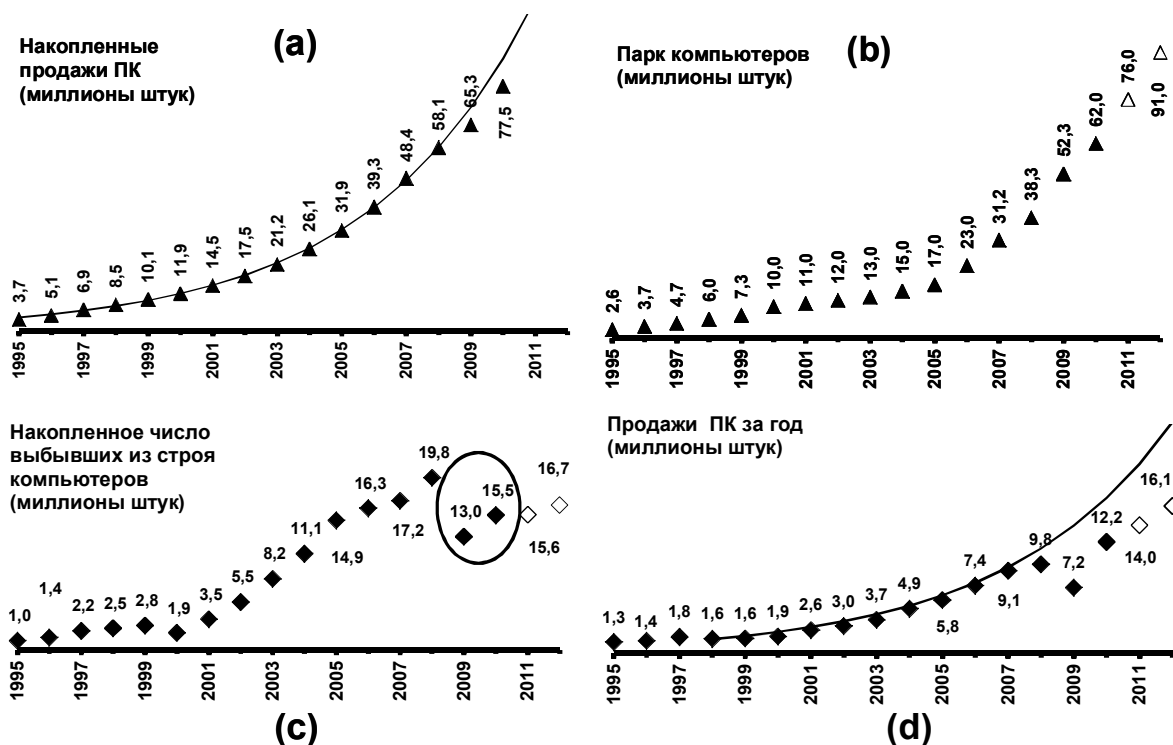


Рис 1. Показатели продаж и парк ПК в РФ в 1995–2010 гг.

Указанные расхождения побуждают нас построить модель и составить прогнозы компьютерного рынка и компьютерного парка Российской Федерации до 2012 г. При построении модели мы опираемся на подход, который учитывает функцию распределения вероятности времени покупки «на замену»¹. Действительно, часть компьютерного парка ежегодно устаревает и нуждается в замене. Гипотеза, которая лежит в основе моделей покупок «на замену», состоит в том, что время замены («возраст замены») есть случайная величина, которая подчиняется распределению вероятностей, аналогичному распределению вероятности поломки продукта. В основе расчетов при этом лежит уравнение теории восстановления

$$\frac{dN}{dt} = q(t) - r(t) = q(t) - \int_0^t q(\theta)g(t-\theta)d\theta, \quad (2)$$

где $N(t)$ — парк компьютеров, $q(t) = \frac{dQ}{dt}$ — продажи ПК в единицу времени (в «штуках»),

$r(t) = \frac{dR}{dt}$ — продажи ПК «на замену» в единицу времени, $g(t)$ — плотность распределения времени жизни ПК.

Разумеется, распределение срока службы компьютеров меняется от года к году. Например, в год финансового кризиса фирма может отложить замену служебных компьютеров своих сотрудников, а частные пользователи — повременить с покупкой нового компьютера. Однако у нас недостаточно данных, чтобы учесть такие флуктуации. Поэтому мы считаем функцию отклика процесса устаревания *стационарной* во времени. Ошибки, возникающие в результате та-

¹ Kamakura W.A., Balasubramanian S.K. Long-term Forecasting with Innovation Diffusion Models: The Impact of Replacement Purchases // Journal of Forecasting. 1988. Vol. 6. P. 1–13.

кого упрощения, не мешают моделировать динамику компьютерного парка РФ с достаточной для наших целей точностью.

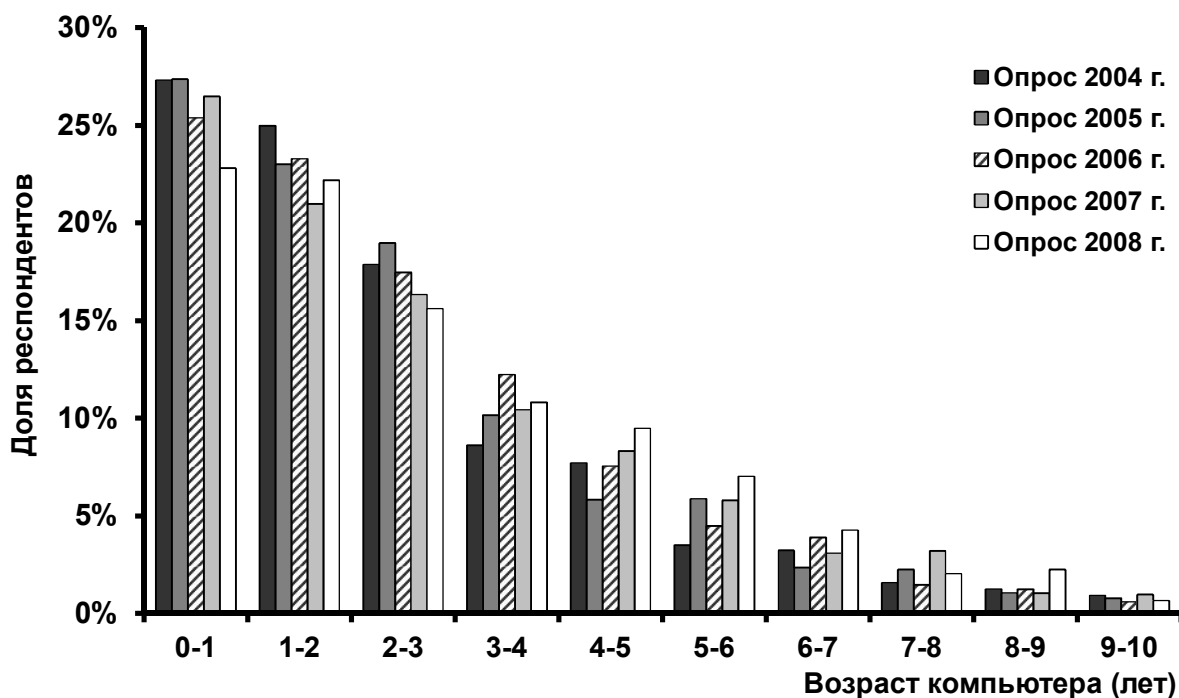


Рис. 2. Итоги пяти онлайн-опросов о возрасте используемого компьютера.

Чтобы определить неизвестную нам функцию $g(t)$, в конце 2004 года мы провели онлайн-опрос о сроке службы компьютера на Интернет-портале Rambler. Основной вопрос звучал так: «В каком году был выпущен компьютер, с которого вы сейчас вышли в Интернет?» Вычитая из года проведения опроса год, который назвал респондент, мы вычисляли возраст компьютера. В ходе опроса 56 тысяч респондентов смогли сообщить дату выпуска своих компьютеров. Безусловно, такой метод проведения опроса нельзя назвать репрезентативным, поскольку мы опрашивали пользователей только тех компьютеров, которые подключены к Интернету, к тому же выборка проводится путем самовыдвижения. К нашему удивлению, сравнение полученных нами данных с результатами дневникового исследования, проведенного компанией КОМКОН, в котором панелистов спрашивали, как долго они пользуются одним компьютером, показали практическую идентичность распределений.

Мы учитывали приобретение только новых компьютеров, однако, как показали опросы, доля компьютеров, приобретенных поддержанными, невелика.

В дальнейшем аналогичный опрос мы повторяли ежегодно в 2005–2008 гг. В каждом случае число респондентов измерялось десятками тысяч. Как показывает рис. 2, итоги опросов за различные годы весьма близки, существенно отличаются лишь ответы за 2008 г., когда доля респондентов, купивших ПК непосредственно в год опроса, была заметно меньше, чем в предыдущие годы. Мы связываем это с резким падением продаж ПК в России в IV квартале, которое (в натуральном выражении) составило 38 %. Мы исключили результаты опросов за 2008 год из процедуры определения распределения срока службы компьютера, поскольку использование таких данных требовало бы знания точной формы продаж $q(t)$ в период их быстрых изменений.

Обозначим $\Delta R(t)$ — количество ПК, вышедших из строя между t и $t + \Delta t$. Эта величина зависит от их возраста компьютеров, и, таким образом, от истории продаж $q(t)$

$$\Delta R(t) = \int_t^{t+\Delta t} \frac{dL(T-\theta)}{d\theta} \cdot q(\theta) d\theta, \quad (3)$$

где $L(t)$ — т.н. «время дожития» или функция надежности устройства, представляющая собой вероятность того, что устройство сохранит работоспособность к моменту t ; $q(t) = \frac{dQ(t)}{dt}$ — интенсивность продаж.

При помощи вероятностных рассуждений нетрудно показать, что модель для определения функции надежности $L(t)$ по данным опроса имеет вид

$$\Delta v_i = c \cdot \int_{t_{i-1}}^{t_i} L(T - \theta) \cdot q(\theta) d\theta, \quad (4)$$

где i — год (1998, ... 2008); t_i — момент окончания этого года; T — момент проведения опроса; Δv_i — теоретическое значение количества респондентов, ответивших, что компьютер, которым они пользуются, заполняя нашу анкету, приобретен новым в i -м году; c — неизвестная нам постоянная, равная доле наших респондентов среди владельцев ПК.

Если $L(t)$ и $q(t)$ между t_i и t_{i+1} меняются медленно, то для расчета Δv_i можно использовать приближение

$$\Delta v_i \approx c \cdot L(T - t_i + \delta) \cdot \Delta Q_i, \quad (5)$$

где ΔQ_i — продажи за год i (или иной период измерения) в натуральном выражении (штуках); $t_i - \delta$ — середина года опроса (данные за 2004 г. относим к середине 2004 г. и т.п.).

Для описания функции надежности в наших предыдущих работах¹ мы использовали двухпараметрическую усеченную логистическую функцию

$$L(x) \equiv L(x; \mu, \tau) = \frac{1 + \exp(-\mu\tau)}{1 + \exp(\mu(x - \tau))} \quad (6)$$

с параметрами положения τ и масштаба μ . Однако при экспоненциально растущих продажах (которые и имели место в 1999–2007 гг.)

$$q(t) = \alpha \exp \beta t \quad (7)$$

наиболее удобным видом функции надежности оказывается гамма-распределение $\gamma(t; \mu, z)$ с параметрами μ и z и плотностью

$$g(t) = \frac{d\gamma}{dt} = \mu \frac{(\mu t)^{z-1}}{\Gamma(z)} e^{-\mu t}. \quad (8)$$

Подставляя $L(t) = \gamma(t; \mu, z)$ и $q(t)$ вида (7) в формулу (4), находим формулу количества респондентов

$$\Delta v_i = c \cdot e^{\beta T} (f(T - t_{i-1}) - f(T - t_i)), \quad (9)$$

где

$$f(a) = \frac{1}{\beta} \left\{ 1 - [1 - \gamma(a; z, \mu)] e^{-\beta a} - \frac{\mu^z}{(\beta + \mu)^z} \gamma(a; z, \beta + \mu) \right\}. \quad (10)$$

¹ Делицын Л.Л. Совместное прогнозирование компьютерного рынка и парка компьютеров при помощи модели распространения нововведения // Труды вольного экономического общества России. 2010. Т. 130. С. 421–429.

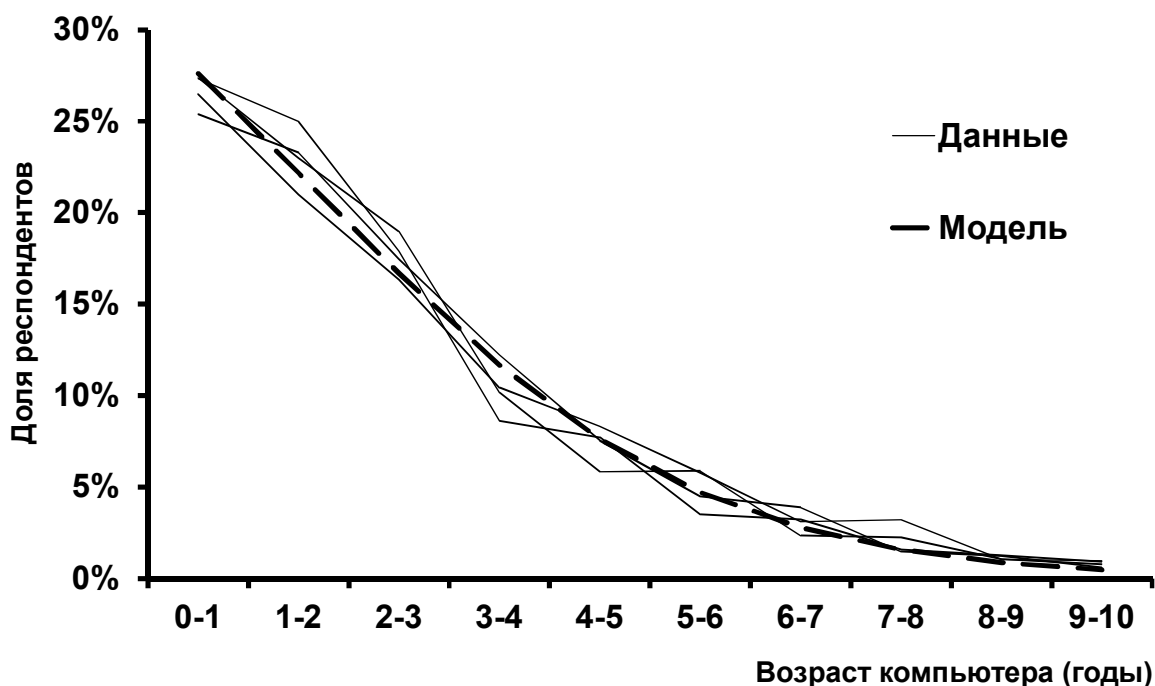


Рис. 3. Итоги моделирования данных опросов о возрасте компьютера.

Поскольку расчет гамма-распределения $\gamma(t; \mu, z)$ является стандартной функцией большинства вычислительных пакетов (включая офисный пакет MS Excel), в случае экспоненциально растущего рынка величину (4) можно рассчитать с высокой точностью, не используя численное интегрирование. Для усеченного логистического распределения (6) также можно выразить число респондентов при помощи специальных функций, однако итоговая формула имеет еще более громоздкий вид. Поэтому в данной работе мы используем гамма-распределение. В более общем случае, когда функция продаж имеет произвольную форму, или при желании работать с поквартальными продажами, уравнение (4) упростить нельзя, и следует интегрировать численно.

Для определения параметров μ , z и c мы использовали метод максимального правдоподобия. На рис. 3 тонкими линиями изображены данные опросов, жирной пунктирной линией — итоги расчетов при помощи формул (9–10) и оптимального набора параметров $\mu = 0,56$, $z = 3,4$. Параметр c различен для каждого года, поскольку число респондентов различалось, в дальнейших вычислениях он не требуется. Средний срок службы компьютера при таком наборе параметров составляет 6,12 года, а кажущаяся задержка между кривыми первых и повторных продаж при темпах прироста рынка в 21 % составляет 5,18 года. Плотность гамма-распределения (8), соответствующая указанным значениям параметров, изображена жирной пунктирной линией на рис. 4 и весьма близка к плотности усеченного логистического распределения, использованного нами ранее. Заметим, что при росте рынка с постоянными темпами (7), который имел место в 1999–2007 гг., любая из плотностей распределения на рис. 4 продуцирует одинаковую кривую повторных продаж (3). Сделанное замечание обосновывает необходимость измерения плотности распределения при помощи опросов, поскольку в условиях экспоненциального роста рынка найти $g(t)$ при помощи известных продаж и парка ПК по формуле (2) невозможно, но для вычисления парка при более сложной зависимости $q(t)$ при помощи (2) потребуется знать точную форму $g(t)$.

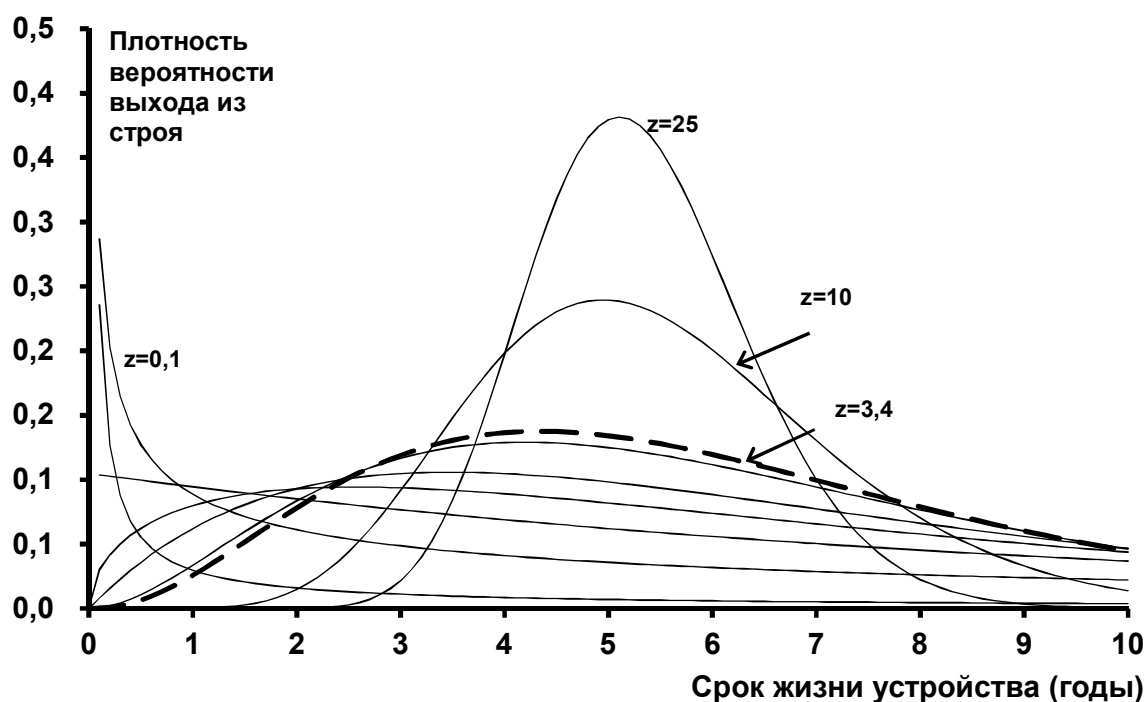


Рис. 4. Теоретические плотности распределения срока службы компьютера.

В 2004–2008 гг. мы проводили опросы о моменте покупки компьютера в конце года, в декабре. Кроме того, дополнительные опросы были проведены в весенние, летние и осенние месяцы, при этом респондентам был задан вопрос о том, в каком месяце они приобрели ПК, что помогло нам определить сезонные множители. Здесь мы ограничимся замечанием о том, что при определении срока службы компьютера предлагаемым нами методом важен месяц проведения опроса, поскольку сезонные колебания продаж весьма сильны, и существуют задержки между покупкой компьютера и подключением его к Интернету, а также между “поставкой” компьютера (которые подсчитывают маркетинговые агентства) и непосредственно продаж. Дальнейшие детали обсуждаются в специальной работе¹.

Нами рассчитаны несколько сценариев роста компьютерного рынка и сопутствующего роста парка компьютеров, при этом проанализированы несколько опубликованных стратегий². В качестве наиболее вероятного сценария роста компьютерного парка нами был предложен представленный на рис. 5 сценарий, основанный на опубликованных прогнозах продаж с темпами прироста рынка персональных компьютеров в 15 % в 2011–12 гг.³. В этом случае уровень продаж в 2012 г. должен составить чрезвычайно высокую величину – 16 миллионов компьютеров, указанную на рис. 1(d). Мы намеренно используем столь высокое значение, чтобы подчеркнуть, что оно не смягчает диапазон расхождений наших оценок с опубликованными прогнозами МЭР. Отметим, что даже грубые ошибки в оценке продаж компьютеров в отдельные годы не повлекут за собой большие отклонения в прогнозах компьютерного парка, поскольку величина парка (в грубом приближении) интегрирует продажи за пять лет.

Мы полагаем, что совместное моделирование данных о продажах компьютеров в РФ и размере компьютерного парка позволяет получать более реалистичные прогнозы обеих величин, избежать грубых рассогласований, подобных продемонстрированным на рис. 1(c), а также исследовать влияние регулирования рынка ПК на парк компьютеров.

¹ Делицын Л.Л. Анализ итогов совместного прогнозирования компьютерного рынка и парка компьютеров РФ на основе новых статистических данных // Труды всерос. конф. «Экономика и инновации в промышленности» / Под ред. проф. А.В. Бабкина. СПб.: Изд-во Политехнического ун-та, 2010. С. 18–25.

² Кузовкова Т.А., Тимошенко Л.С. Анализ и прогнозирование развития инфокоммуникаций. М.: Горячая линия – Телеком, 2009; Министерство Экономического развития. Прогноз социально-экономического...

³ Делицын Л.Л. Совместное прогнозирование...

Парк компьютеров РФ: сценарий F2010



Рис. 5. Компьютерный парк РФ: модель и прогнозы.

Опубликованный весной 2010 г. Министерством экономического развития прогноз парка ПК объемом в 63 млн штук, подтвержденный позже в публикациях Минкомсвязи оценки в 62 млн устройств более, чем на 40 % превосходит размеры, которые мы получили путем расчетов по модели (2). Расхождения прогноза Минэкономразвития с нашим прогнозом 'F' на 2011 г. являются почти двукратными. По мнению министерства, «к 2012 году по сравнению с 2008 годом количество персональных компьютеров возрастет более чем в 2 раза, достигнув 97 миллионов»¹. Оптимистический и консервативный прогнозы Минэкономразвития изображены двумя серыми линиями в правом верхнем углу рис. 5.

Безусловно, одной из причин расхождений может быть грубая ошибка в оценке объема компьютерного рынка агентствами IDC, Gartner и IT-Research, в частности недоучет т.н. «самосбора». Однако методика IT-research использует подсчет количества ввезенных в РФ жестких дисков, которые не производятся внутри страны и не могут быть собраны самостоятельно. Более того, оценки IT-Research, по утверждению компании, включают и собранные самостоятельно компьютеры. Расхождения в оценках между агентствами невелики, и использование максимальной оценки вместо средней не привело к существенному снижению расхождений в размере парка ПК. Наконец, причиной различий может быть ошибка в определении нами плотности вероятности выхода из строя компьютера. Средний срок службы ПК может оказаться выше, чем оценка в 6,12 года, полученная нами путем моделирования ответов пользователей лишь тех компьютеров, которые используются для выхода в Интернет. Не исключено, что компьютеры, не подключенные к Интернету, а используемые в производстве или для ведения личных финансов и набора текстов, используются в *среднем* значительно дольше. Несмотря на сделанные замечания относительно численных значений, наша модель обладает преимуществами внутренней согласованности, и не допускает расхождений между продажами и парком ПК, продемонстрированных на рис. 1.

¹ Министерство экономического развития. Прогноз социально-экономического...

Опросы в мобильном Интернете: кого и как можно опрашивать

Зарин Михаил Константинович
ООО «Мобильные опросы»

В сентябре 2010 года аналитиками компании «Мобильные опросы» было проведено исследование, посвящённое использованию жителями Москвы и Санкт-Петербурга мобильного Интернета. Исследование было реализовано методом телефонного опроса по мобильному телефону. Случайная выборка контролировалась по половозрастным квотам и составила 1953 человека. Цель исследования — составить портрет пользователя мобильного Интернета, понять, насколько существенен этот сегмент для привлечения к участию в мобильных опросах.

Сомнения в эффективности существующих методик телефонных опросов

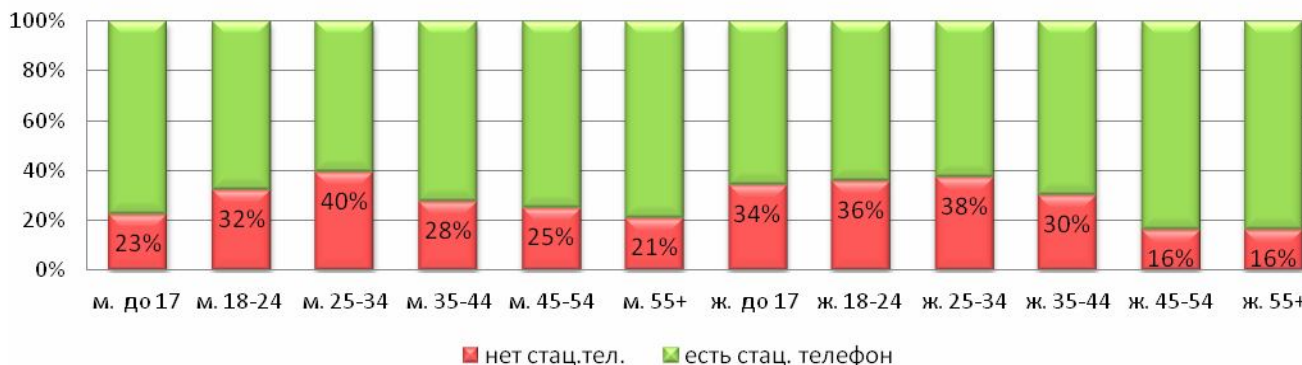
Интерес к опросам, использующим мобильный Интернет, возник не случайно. Известно, что при проведении социологических и маркетинговых исследований зачастую используются стандартные квартирные телефонные опросы. Но так ли они эффективны в достижении квот, как кажется, если, по данным проведённого исследования, у трети владельцев мобильных вообще нет стационарного телефона (см. Диаграмму 1)?

Диаграмма 1. Есть ли у вас дома стационарный телефон?



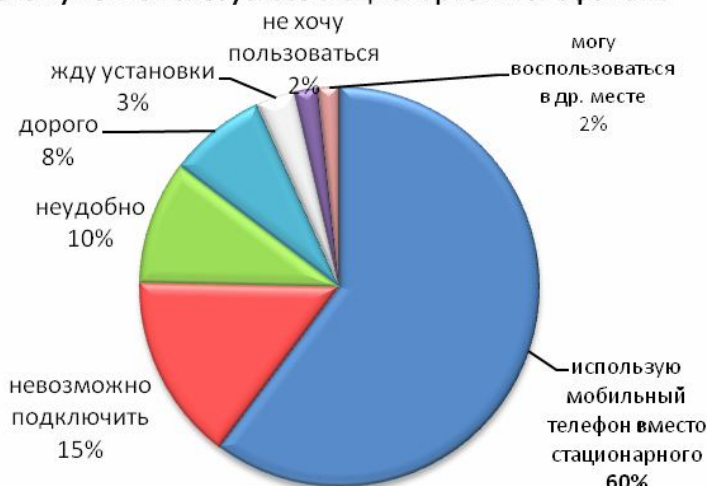
Как видно из диаграммы 2, 40 % мужчин 25–34 лет, имеют мобильный телефон, но не имеют стационарного, а значит, не могут быть охвачены телефонным анкетированием, проводимым по базе городских номеров. То же относится и к остальным демографическим группам респондентов.

Диаграмма 2. Соотношение абонентов и не абонентов стационарной связи в разрезе половозрастных характеристик.



Как оказалось, при классических телефонных опросах исследователи упускают довольно большой пласт городского населения, которые не пользуются (не подключили или, наоборот, отказались от него) стационарным телефоном. Большинство респондентов (60 %) используют мобильный телефон вместо домашнего, 15 % не имеют технической возможности провести телефон в свой дом, для 10 % опрошенных это дорого, для 8 % — неудобно.

Диаграмма 3.
Почему Вы не пользуетесь стационарным телефоном?



Вывод: при опросах населения необходимо использовать методы, позволяющие достичь контактов со всеми группами населения, а обычный «телефонник», как установлено по итогам данного исследования, не может нам в этом помочь, равно как и дискредитировавшие себя с точки зрения контролируемости исполнения методы квартирных и личных опросов. Но кроме телефонного анкетирования по базам мобильных номеров, есть новый метод удалённого опроса, отличающийся высокой оперативностью и возможностью получить репрезентативные данные — это опросы с помощью мобильного Интернета.

Однако сначала нужно изучить самих пользователей мобильного Интернета. Кто они? Так ли их мало среди нас, как кажется на первый взгляд?

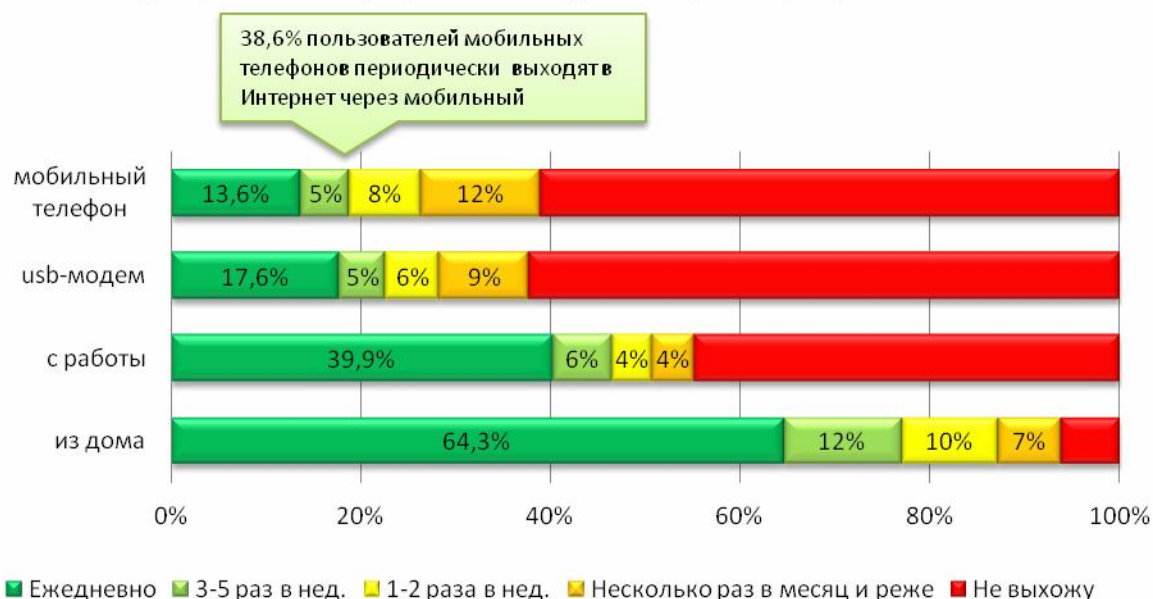
Как показывают результаты опроса, большая часть владельцев мобильных телефонов (66 % от всех опрошенных) так или иначе пользуется Интернетом.

Диаграмма 4.
Использование Интернета владельцами мобильных телефонов.



А 38,6 % (!!!) Интернет-пользователей периодически (от пары раз в месяц до пары раз в день) выходят в глобальную паутину именно через мобильный телефон!

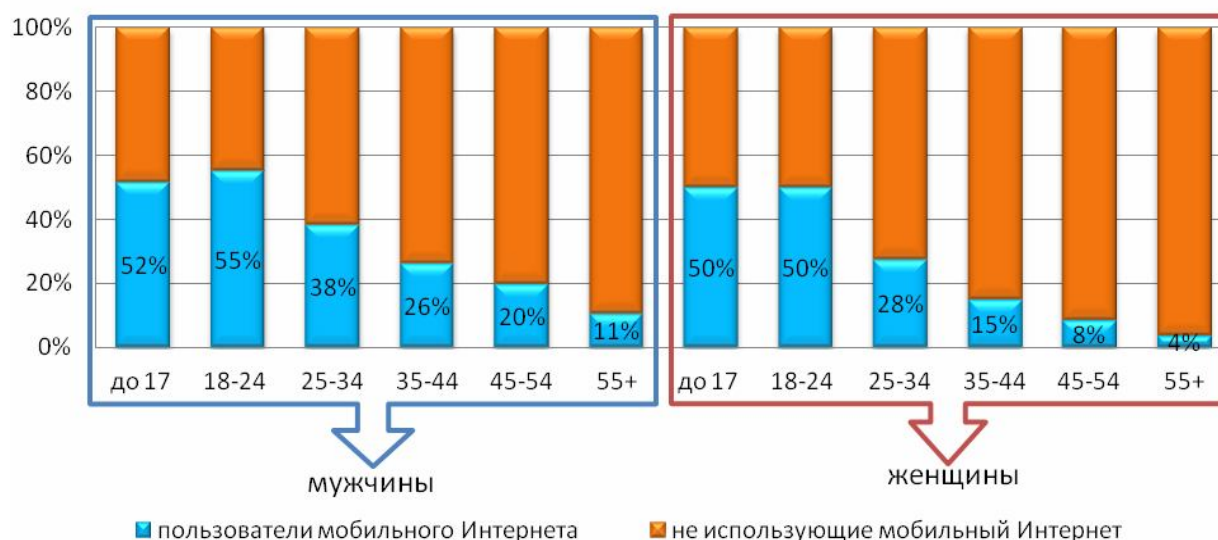
Диаграмма 5. Регулярность выхода в Интернет через разные каналы.



Кто же эти люди?

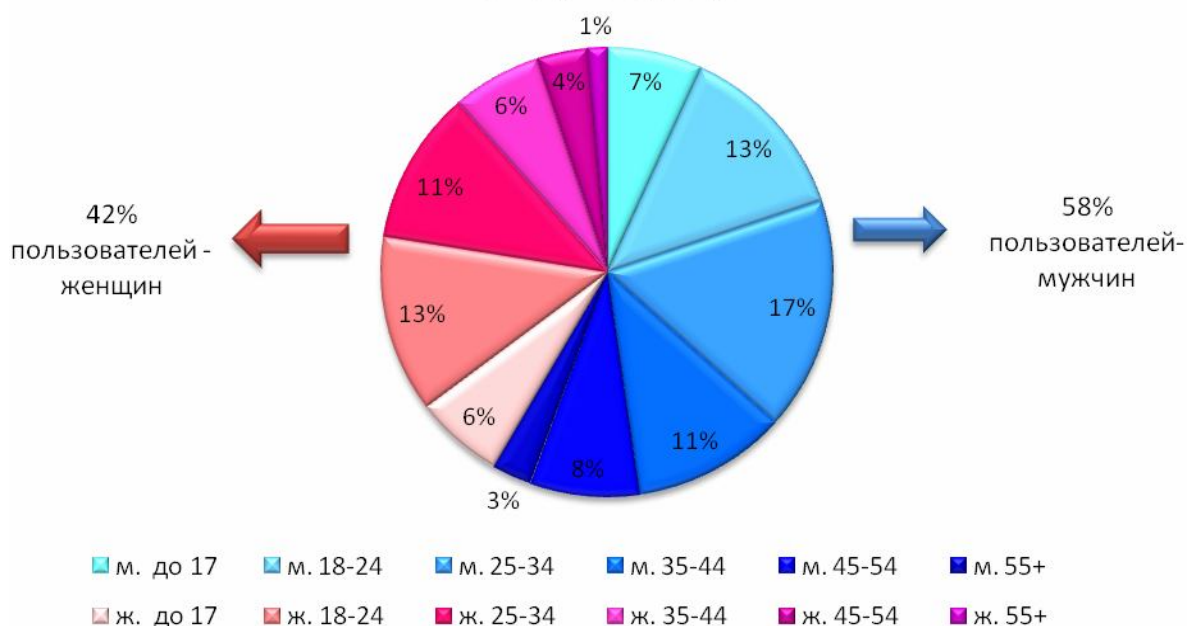
Среди владельцев мобильных телефонов мобильным Интернетом пользуются в основном мужчины до 54 лет и женщины до 44 лет; после этого возрастного предела число абонентов мобильного Интернета стремительно снижается (см. диаграмму 6).

Диаграмма 6. Доля пользователей мобильного интернета среди владельцев мобильных в различных демографических группах.



В самом сегменте пользователей мобильного Интернета, как видно из диаграммы 7, преобладают мужчины 18–54 лет (41 % пользователей) и женщины 25–44 лет (24 %). Но нельзя исключать и другие возрастные категории, кроме совсем уж небольшого процента пользователей старше 55 лет.

Диаграмма 7. Распределение пользователей мобильного интернета по полу и возрасту.



Индекс совпадения социально-демографических характеристик (Affinity Index)

Индекс совпадения характеристик (Affinity Index) — индекс совпадения пользователей мобильного Интернета и всех владельцев мобильных телефонов. Значение индекса больше 100 показывает, что в этой демографической группе находится большее число пользователей мобильного Интернета, чем в других группах. Значение индекса меньше 90 показывает, что в этой демографической группе находится меньшее число пользователей мобильного интернета, чем в других группах.

Очевидно, пользователями мобильного Интернета преимущественно являются студенты. Однако большая доля пользователей есть и среди наёмных сотрудников коммерческих и некоммерческих организаций, а так же владельцев собственного бизнеса и индивидуальных предпринимателей.

Диаграмма 8. Затраты на мобильный телефон в месяц



Наиболее открыты к пользованию мобильным Интернетом люди, занятые в СМИ, ИТ, сфере услуг и здравоохранении. Пользоваться мобильным Интернетом типичнее для владельцев телефонов с высоким и средним доходами. Об этом косвенно говорят и данные, отражённые на диаграмме 8. Потребители мобильного Интернета тратят денег на телефон в месяц больше по сравнению с теми, кто его не использует: почти половина (47 %) потребителей мобильного Интернета платит в месяц от 1000 рублей и больше, в то время как такие же расходы принимает на себя только 34 % тех, кто не пользуется мобильным Интернетом. К сожалению, это отражение реальной ситуации на рынке мобильных услуг РФ: выход в Глобальную Сеть всё ещё довольно дорог для большинства потребителей.

Состав мобильного Интернет-трафика

Основной целью выхода в Интернет с помощью мобильного телефона респонденты называют поиск информации (64 % выходят в Сеть для этого), 52 % просматривают почту, всеобъемлющие, на первый взгляд, социальные сети интересуют только 42 % абонентов мобильного Интернета, 38 % смотрят новости, 20 % проверяют обновления блогов.

Диаграмма 9.
Цели выхода в Интернет через мобильный телефон

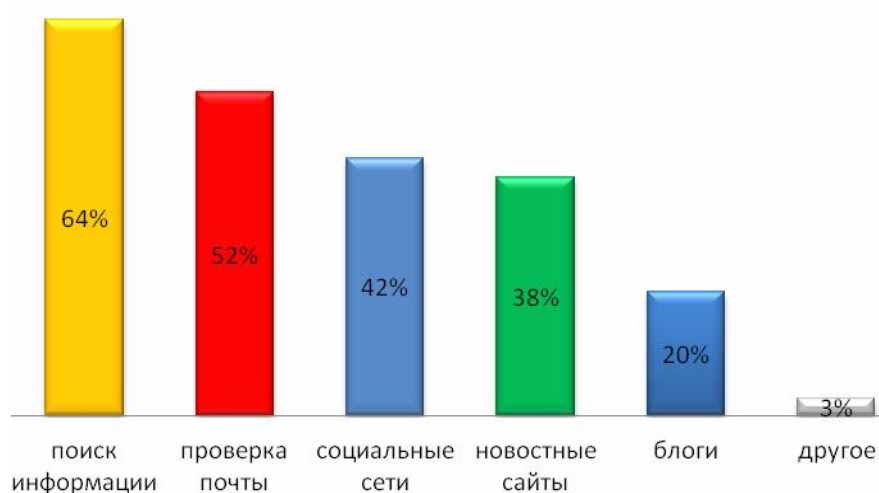
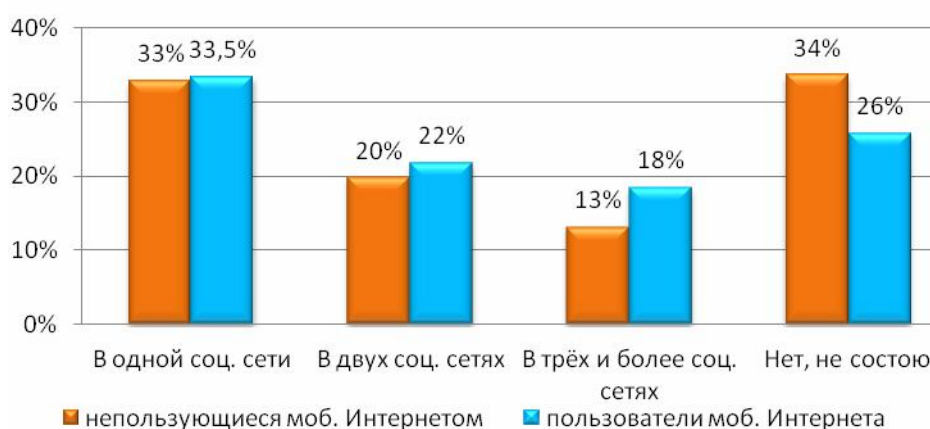


Диаграмма 10. Состоите ли вы в социальных сетях. Если да, в сколько?

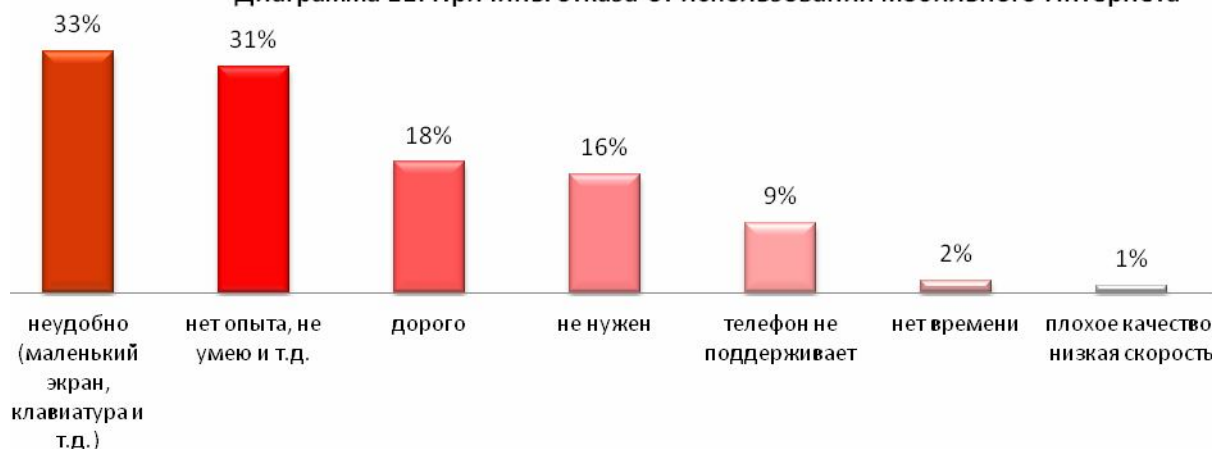


Пользователи мобильного Интернета немного более социально активны в Сети по сравнению с не пользующимися этой услугой: 18 % из них состоит в трех и более социальных сетях, в то время как такой же показатель среди всех владельцев мобильных, не выходящих в Сеть с помощью телефона, составляет 13 % (см. Диаграмму 10).

Основания для отказа от использования Мобильного Интернета

Почему же не все пользователи мобильных телефонов используют услугу Мобильного Интернета? Оказывается, для 33 % это неудобно в силу различных причин, связанных с телефоном (маленький экран, клавиатура т.д.), 31 % просто не умеет выходить в Интернет с мобильного устройства, для 18 % это дорого, 16 % эта функция не нужна и т.д.

Диаграмма 11. Причины отказа от использования мобильного Интернета



Выводы

Таким образом, нам удалось нарисовать примерный портрет основного потребителя Мобильного Интернета: преимущественно мужчины до 44 и женщины до 35 лет, студенты, владельцы собственного бизнеса, наёмные сотрудники, работающие в СМИ, ИТ, сфере услуг и здравоохранении со средним и высоким доходами. Это социально активные люди, которые пользуются в Интернете, в первую очередь, поисковиками, почтовыми сайтами, а уже заодно с ними и социальными сетями, просматривая так же и новостные сайты.

Этот сегмент более чем реален, и по мере дальнейшего проникновения Интернета в жизнь общества будет расти, особенно учитывая постоянное снижение цен на мобильный доступ в Интернет и обновление парка мобильных телефонов. Учитывая, что примерно пятая часть отказов от использования Мобильного Интернета происходит из-за дороговизны услуги, можно смело предположить, что при снижении тарифов количество пользователей Мобильного Интернета увеличится. То же самое можно сказать и о тех, кто отвергает использование мобильного Интернета из-за неудобства использования или отсутствия этой функции в телефоне: нельзя игнорировать отчётливый тренд на рынке мобильных телефонов в сторону выпуска всё более многофункциональных устройств, использование Интернета в которых скоро станет второй по использованию опцией после звонков.

Всё вышесказанное говорит о невозможности и далее не замечать стремительно формирующийся сегмент потребителей мобильного Интернета. Данный метод более оперативно, чем существующие методики опросов, получать репрезентативные результаты.

Метод множественного восстановления данных

Кутлалиев Асхат Хасянович

*Международный институт маркетинговых
и социальных исследований «ГфК-Русь»*

Введение

Наиболее распространенным инструментом социальных исследований является массовый опрос. При этом крайне редко можно встретить количественное исследование, база которого не содержит пропущенных данных. Неполнота массива данных происходит вследствие целого комплекса причин: недостижимости отдельных респондентов, необходимых для поддержания репрезентативности в выборке, пропущенных ответов некоторых респондентов на отдельные вопросы вследствие усталости, недостатка знаний, социальной чувствительности вопросов.

Кроме того, с распространением компьютерных технологий сбора информации, особенно CATI и CAWI, исследователи всё чаще сталкиваются с ростом доли незавершенных интервью, особенно при проведении сложных комплексных опросов. Всё вышеперечисленные причины ведут к смещению выборочного массива, неполноте получаемой информации, недостоверности результатов, что является достаточно серьезными проблемами. Исследователи вынуждены искать способы учета и работы с пропущенными данными.

К сожалению, не существует достаточно простых и общепринятых правил как это делать. В каких случаях можно не обращать внимания на пропуски информации, в каких случаях исключать из дальнейшего анализа наблюдения, содержащие пропуски данных и в каких – стараться тем или иным способом заполнять пропуски. Например, часть пропусков заполняется в процессе редактирования данных после их сбора и ввода. Но таким образом можно достоверно заполнять пропуски лишь в тех вопросах, которые могут быть логически проверены ответами на другие вопросы. При больших выборках с небольшим (не более 5 %) количеством наблюдений с пропущенными данными неполные наблюдения (даже при наличии всего лишь одного пропуска) часто исключают из анализа. Наблюдения с пропусками в данных в широко распространенном статистическом пакете SPSS удаляются по умолчанию в большинстве процедур регрессионного и многомерного анализа. Даже при большей доли наблюдений с пропусками в данных многие исследователи предпочитают исключать наблюдения из анализа целиком, полагая, что заполнение пропусков ведет к искажению регрессионных коэффициентов, мер связи и смещению статистических оценок. И для таких предпочтений в отдельных случаях есть основания. Но более правильным кажется альтернативный путь: выявить природу пропусков, попытаться восстановить пропущенную информацию и сравнить результаты с теми, которые получены при анализе массива полных наблюдений.

В ряде случаев процедура восстановления пропусков в данных является обязательным элементом анализа, например, если исследователь сознательно планирует контролируемые пропуски в данных. Такой дизайн исследования может включать в себя опрос по отдельным блокам вопросов только части выборки, опрос различных групп людей из одного выборочного массива с помощью частично совпадающих анкет, охват мониторингом различных групп респондентов в разные периоды времени и/или разной тематикой.

Типы пропусков в данных и методы их восстановления

Согласно классификации Литтла и Рубина¹, в данных могут встречаться:

- 1) полностью случайные пропуски (missing completely at random (MCAR));
- 2) случайные пропуски (missing at random (MAR));
- 3) неигнорируемые пропуски (non-ignorable missingness).

Последний тип пропусков, наиболее сложный для анализа и восстановления, мы в данной работе не рассматриваем, а сосредоточиваемся на первых двух типах пропусков данных.

С полностью случайными пропусками (MCAR) мы имеем дело тогда, когда пропуски случайно распределены в массиве данных по всем переменным. Наличие MCAR можно проверить статистически t-тестом или хи-квадрат тестом. В модуле SPSS Missing Values Analysis (MVA) есть опция Little's MCAR test, которая на базе статистики хи-квадрат проверяет данные на MCAR. Если в тесте наблюдается *незначимый* уровень критерия, то мы имеем дело с полностью случайными пропусками данных.

Случайные пропуски (MAR) данных встречаются тогда, когда пропуски в массиве данных случайно распределены не по всем переменным, а только внутри каких-либо определенных подгрупп переменных. Например, пропуски чаще встречаются среди ответов мужчин, чем ответов женщин, но внутри подгрупп распределены случайны. Такое распределение пропусков в данных случается гораздо чаще, чем MCAR.

Для борьбы с этими двумя видами пропусков применяют восемь основных классов методов:

- 1) анализ полных наблюдений (listwise deletion);
- 2) методы, использующие доступную информацию (pairwise deletion);

¹ Little R.J.A., Rubin D.B. Statistical Analysis with Missing Data. New York: John Wiley & Sons, 1987.

- 3) подстановка среднего по выборке (mean substitution);
- 4) метод хот-дек (hot deck);
- 5) регрессионный анализ (regression);
- 6) оценка с помощью максимизации правдоподобия (maximum likelihood estimation);
- 7) подстановка с помощью факторного анализа (factor analysis substitution);
- 8) модель множественного восстановления данных (multiple imputations method).

Два первых метода широко распространены в исследовательской практике.

Смысл первого метода заключается в удалении всех респондентов с пропущенными данными (listwise deletion). При вычислении сложных моделей с большим числом переменных этот метод может привести к большому объему исключенных наблюдений, в ряде случаев до двух третей исходного массива данных. Во втором методе переменные для наблюдений с пропущенными значениями не включаются в проводимый анализ по необходимости (pairwise deletion). Например, корреляции между каждой парой переменных вычисляются из наблюдений, которые для этих переменных содержат допустимые значения, что позволяет использовать больший объем доступной информации по сравнению с анализом полных наблюдений. Применение этих способов методически правомерно, если пропуски в данных распределены полностью случайно (MCAR).

Методы с 3-го по 7-ой используют принцип однократной подстановки восстановленных тем или иным способом данных и могут использоваться, если пропуски распределены случайно (MAR). В основе методов лежит принцип вычисления и подстановки взамен каждого пропущенного значения одного нового значения. Кратко рассмотрим эти методы.

Суть метода **подстановки среднего по выборке** понятна из его названия: вместо пропусков подставляется среднее значение по данной переменной. Метод видоизменяет изначальное распределение, делая его более сконцентрированным около среднего значения и уменьшая дисперсию. Как следствие, из-за подстановки среднего по выборке систематически недооцениваются ковариации переменных. Этот метод дает лучшие результаты, чем методы, основанные на исключении наблюдений, однако он также ведет к смещенным результатам.

Метод хот-дек (hot deck) имеет ряд модификаций. Наиболее простой вариант — сортировка респондентов по ключевым переменным, тогда респонденты со схожими ответами находятся рядом друг с другом. В качестве ключевых чаще всего выступают социально-демографические переменные, но можно использовать и другие переменные, имеющие корреляции с переменной с пропущенными данными. При восстановлении пропущенные значения переменной заимствуются из предыдущего наблюдения. В отдельных модификациях хот-дек респонденты сортируются для каждого признака при помощи разных наборов переменных для предупреждения взаимозависимостей между восстанавливаемыми значениями. Наиболее продвинутой модификацией считается хот-дек со случайным отбором значений из подгруппы схожих респондентов. Респонденты делятся на подгруппы по ключевым переменным. Внутри каждой подгруппы данные для замены пропущенных ответов отбираются случайным образом, но с тем же распределением, что и все ответы в подгруппе. Главным недостатком метода хот-дек является наличие взаимосвязей между восстановленными значениями и занижение значений дисперсии. Снизить уровень зависимости до незначимых величин возможно лишь при очень большом числе подгрупп, что подразумевает большие объемы выборок. Тем не менее, метод широко применяется в работе национальных статистических организаций многих стран, включая страны Евросоюза, Россию и США.

Регрессионный анализ. В зависимости от типа данных используются либо множественная линейная, либо логистическая регрессия. На базе наблюдений, не содержащих пропущенных данных, вычисляются коэффициенты регрессии, и далее с их помощью восстанавливается пропущенное значение зависимой переменной. Если не говорить об обычных проблемах регрессии, таких как мультиколлинеарность, гомоскедастичность и т.д., то можно обозначить две проблемы, связанные с её использованием для восстановления пропущенных данных. Во-первых, из-за самой природы регрессии мы полностью исключаем случайные вариации. Это, например, проявляется в том, что при одинаковом наборе значений независимых переменных мы будем подставлять в разные наблюдения одни и те же восстановленные значения. Это ведет к тому,

что при большой доле пропущенных значений становится очень заметным смещение результатов по направлению к средним оценкам. Для борьбы с этим используется метод случайной подстановки, при котором к вычисленному значению прибавляются случайные величины, например, из набора остатков уравнений регрессии на полных наборах данных. Во-вторых, используя в уравнение регрессии слишком большой набор независимых переменных, мы рискуем моделировать шум вместо каких-то осмысленных значений переменных.

Метод **максимизации правдоподобия (maximum likelihood estimation, MLE)** предназначен специально для работы с недостающей информацией на больших выборках. Другое название метода, под которым он известен пользователям SPSS, — Expectation Maximization algorithm, алгоритм максимизации математического ожидания.

Идея метода основана на допущении: то, что «случилось» в исследовании, то и должно было произойти, то есть реализовались события, наиболее вероятные в исследуемой системе и, одновременно, наиболее соответствующие применяемому инструменту исследования. Поэтому, все неизвестные данные, которые мы пытаемся восстановить, надо искать таким образом, чтобы они как можно лучше согласовывались с уже имеющимися данными в базе данных. Тогда оценки пропущенных данных и будут «наиболее правдоподобными».

Процесс итераций состоит из двух шагов. На шаге Expectation вычисляется ожидаемое значение логистической вероятности математического ожидания для всех данных на основе условного распределения пропущенных значений относительно наблюдаемой информации и текущих оценок интересующего нас параметра из предыдущей итерации. На шаге Maximization максимизируется конечная функция для получения новых оценок параметра. После определенного набора итераций этих двух шагов решение с некоторой точностью приходит к локальному максимуму логистической вероятности математического ожидания наблюдаемых значений. Как и в регрессионном анализе, использование метода максимизации правдоподобия приводит к снижению дисперсии и риску смоделировать шум, однако в настоящее время он получил наиболее широкое применение благодаря тому, что достаточно давно получил реализацию в популярных статистических пакетах (в SPSS, начиная с 8-ой версии).

Метод **подстановки значений с помощью факторного анализа** был разработан сравнительно недавно, в самом конце XX века. Первоначально Майклом Веделем и Вагнером Камакурой¹ была предложена модификация алгоритма факторного анализа при наличии большого числа пропущенных данных. Но оказалось, что, используя алгоритм максимизации правдоподобия, можно решить и обратную задачу — подстановку пропущенных значений исходных переменных на базе полученных латентных переменных, т. е. факторов. Этот метод основан на предположении, что все факторы можно вычислить на основе известных, наблюдаемых значений переменных. Метод работает с различными типами данных, что является большим преимуществом. Разработчики метода особенно подчеркивают, что предложенная ими модель подстановки оправдана даже в случае пропуска 50 % данных и дает более достоверные результаты, чем анализ полных наблюдений, метод, использующий доступные данные, и подстановка среднего по выборке.

Метод множественного восстановления данных

Среди методов восстановления пропущенных данных наилучшим на данный момент является метод множественного восстановления (multiple imputations), предложенный в конце 1970-х Дональдом Рубином². От предыдущих этот метод отличает сложность и ресурсоёмкость, что ставит перед практикующими исследователями серьезные методологические трудности, препятствующие его грамотной практической реализации. А публикации на русском языке, посвященные данному методу, можно пересчитать по пальцам одной руки^{3,1}. В первую очередь

¹ Kamakura W.A., Wedel M. Factor Analysis and Missing Data // Journal of Marketing Research. 2000. Vol. 37. No. 4: Nov. P. 490–498.

² Rubin D.B. Inference and Missing Data // Biometrika. 1976. Vol. 63. P. 581–592.

³ Косьяненко А.В. Опыт восстановления пропущенной рыночной информации на основе Байесовского подхода: Препринт WP16/2007/02. М.: ГУ–ВШЭ, 2007.

это вызвано тем, что до недавнего времени в статистическом пакете SPSS, который *de facto* является в России стандартом в области анализа данных в социальных науках, этот метод отсутствовал. Другая причина, сдерживающая применение метода, состоит в том, что в его основу положены редко встречающиеся в практике прикладных социальных исследований байесовский подход и алгоритм Монте-Карло с применением марковских цепей для статистического моделирования. Поэтому закономерен вопрос: стоит ли в данном случае полученный результат затраченных усилий.

Суть метода множественного восстановления данных заключается в том, чтобы одновременно генерировать несколько значений пропущенной величины вместо того, чтобы заменять пропущенную информацию одним значением. Не ограничиваясь однократным вычислением ожидаемого (среднего) значения для пропущенной информации, можно случайным образом подставить значения, вычисленные на основе предсказанного распределения переменной. Для практического использования всех восстановленных таким образом данных генерируются наборы баз данных с различными вариантами подстановок пропущенных данных.

Метод многократной подстановки включает несколько этапов. На первом этапе необходимо выбрать модель, с помощью которой будут предсказываться пропущенные значения. На втором этапе производится несколько вариантов подстановки значений, при этом для каждого варианта генерируется своя база данных с набором восстановленных данных. Завершается алгоритм анализом всех полученных баз данных с подставленными значениями и объединением оценок интересующих нас параметров в общие окончательные показатели (см. Рис.) с помощью алгоритма Рубина².

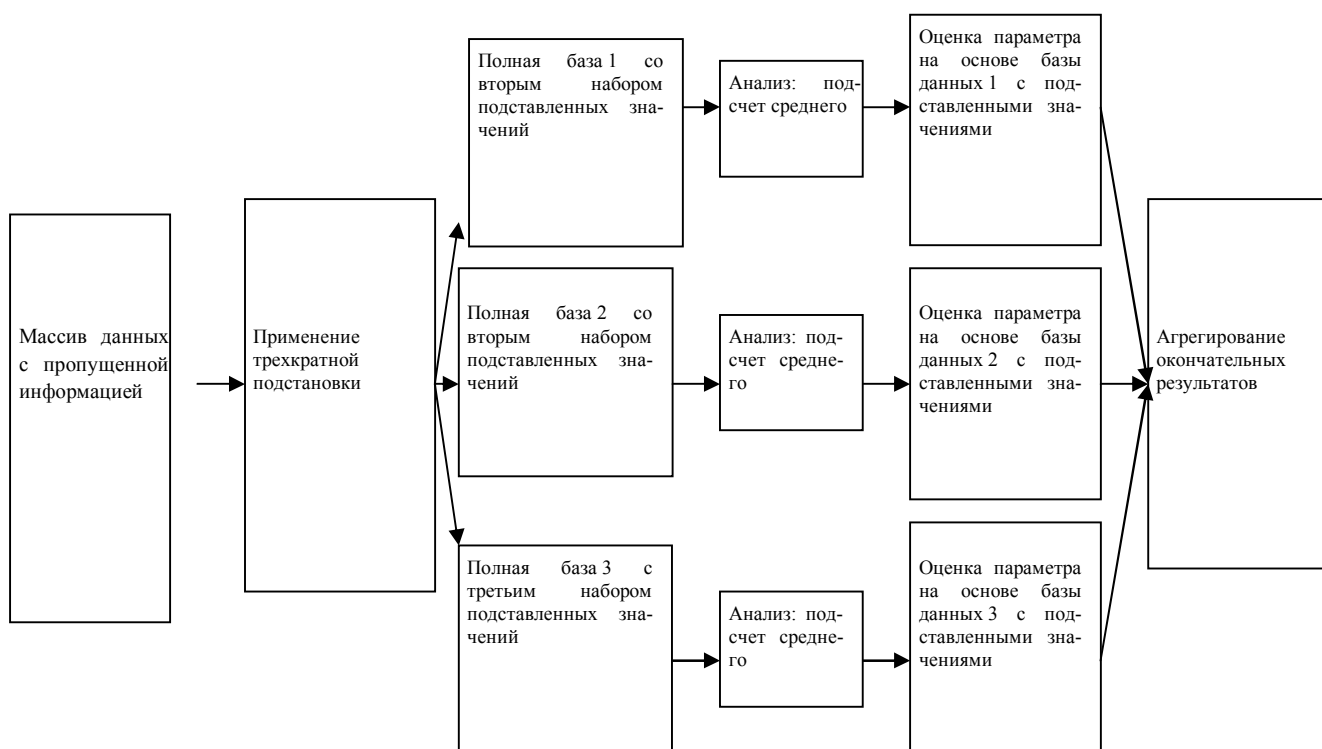


Рис. Схема анализа данных с использованием метода множественного восстановления данных

Каким образом происходит агрегирование данных и оценка их статистических параметров³? Допустим, мы создали m наборов данных с восстановленными пропусками. Для каждого набора мы должны вычислить оценки средних и стандартные ошибки (оценки дисперсий). Обозна-

¹ Зангиева И.К. Решение проблемы неполноты данных массовых опросов // Российская социология завтрашнего дня / Под. ред. Г.В. Иванченко, И.С. Чирикова; факультет социологии ГУ-ВШЭ. М., 2008. С. 84–95.

² Rubin D.B. Multiple Imputation for Nonresponse in Surveys. New York: J. Wiley & Sons, 1987.

³ Schafer J.L. Multiple imputation: A primer // Statistical Methods in Medical Research. 1999. Vol. 8: No. 1. P. 3–15.

чим оценку интересующего нас параметра (например, коэффициента в уравнении регрессии), как $\hat{Q}_j$, полученное из j-ого набора данных ($j=1,2,\dots,m$), а U_j — стандартная ошибка параметра $\hat{Q}_j$. Тогда агрегированная оценка среднего будет равна:

$$\bar{Q} = \frac{1}{m} \sum_{j=1}^m \hat{Q}_j.$$

Для оценки агрегированной стандартной ошибки сначала вычисляется внутригрупповая дисперсия:

$$\bar{U} = \frac{1}{m} \sum_{j=1}^m U_j.$$

и межгрупповая дисперсия:

$$B = \frac{1}{m-1} \sum_{j=1}^m (\hat{Q}_j - \bar{Q})^2.$$

Тогда общая дисперсия будет равна:

$$T = \bar{U} + \left(1 + \frac{1}{m}\right) B.$$

Доверительные интервалы вычисляются стандартно, но с учетом того, что используется значение статистики Стьюдента с df степенями свободы, вычисляемой по формуле:

$$df = (m-1) \left(1 + \frac{m \bar{U}}{(m+1)B}\right)^2.$$

Проверка нулевой гипотезы $Q=0$ выполняется сравнением соотношения $t = \bar{Q}/\sqrt{T}$ с тем же распределением Стьюдента.

Как было доказано разработчиками метода, в большинстве случаев достаточно небольшого числа подстановок — от трех до десяти. Эффективность оценки, основанной на m-ом количестве подстановок, можно вычислить по формуле¹:

$$\varepsilon = \left(1 + \frac{\gamma}{m}\right)^{-1},$$

где γ — доля пропущенной информации в массиве данных. Наглядное представление об эффективности даёт следующая таблица.

Таблица

Эффективность оценки пропущенной информации
методом множественного восстановления данных

m	Эффективность оценки (%)			
	$\gamma=0.1$	$\gamma=0.3$	$\gamma=0.5$	$\gamma=0.7$
3	97	91	86	81
5	98	94	91	88
10	99	97	95	93

Для эффективного вычисления среднего, дисперсии и ковариации (или корреляции), используется алгоритм максимизации правдоподобия (ЕМ), позволяющий учесть все значения массива данных, в том числе и частично пропущенные. Для генерации значений для пропущенной информации обычно используются процедуры, имеющие в своей основе цепи Маркова и метод Монте-Карло.

Метод работает следующим образом. Приращение данных начинается со случайного приписывания значений пропущенной информации, после чего новые оценки параметра извлекаются из предыдущего распределения Байеса, основанного на приписанных данных. Алгоритм под-

¹ Schafer J.L. Analysis of Incomplete Multivariate Data. London: Chapman & Hall, 1997.

становки данных состоит из двух шагов. На первом шаге новые значения для пропущенной информации приписываются исходя из условного распределения наблюдаемых данных и последних приписанных значений параметров. На втором шаге новые значения для параметров симулируются путем извлечения их из последующих распределений Байеса наблюдаемых и последних приписываемых значений для пропущенной информации. Начальные значения параметров для первых подстановок можно получить с помощью алгоритма максимизации правдоподобия. Многократные подстановки производятся путем применения цепей Маркова и метода Монте-Карло. Проведение этой процедуры для массива данных с большим числом переменных может занять много времени.

Метод множественного восстановления данных довольно сложен как в техническом, так и в содержательном плане, поэтому вполне закономерным является вопрос о тех преимуществах, которые этот метод имеет по сравнению с другими. Был выполнен вычислительный эксперимент¹ на 1000 тестовых массивов данных объемом порядка 300 наблюдений в каждом и сгенерированными пропущенными данными. Во всех массивах даны был одинаковый процент пропущенной информации, случайно распределенной по одним и тем же законам распределения. Массивы данных обрабатывались различными способами борьбы с пропущенными значениями: подстановка среднего, подстановка на основе регрессионного анализа, модель максимизации правдоподобия (ЕМ) и модель множественных подстановок. Сравнение метода множественного восстановления данных с другими показало, что он позволяет избежать систематических ошибок в оценке коэффициентов регрессионных моделей, а также превосходит другие методы при большом проценте пропущенных данных. Можно считать, что в настоящее время модель многократных подстановок, с учетом возможности его практического применения (начиная с 17-ой версии SPSS) и качества получаемых результатов, является наилучшим из имеющихся. В отличие от других рассмотренных нами методов, он учитывает неточность подстановок, что приводит к получению более близких к реальности результатов анализа.

Применение метода в практике исследований не должно ограничиваться только вынужденным восстановлением пропущенных данных *post hoc*². При использовании современных компьютерных методов сбора информации появилась необходимость сокращать время интервью для увеличения доли полностью заполненных анкет. В таких случаях приходится сознательно закладывать в дизайн исследования пропуски в данных, например, случайным образом исключая отдельные вопросы или блоки вопросов для разных респондентов, а затем для способов восстановления пропущенной информации использовать метод множественного восстановления данных.

Заключение

Пропущенная информация в массивах данных социологических и маркетинговых исследованиях является скорее правилом, чем исключением. Поэтому обязательным этапом исследования является процедура редактирования данных. Хотя часть информации может быть восстановлена на основе заполненных анкет, большая часть пропусков либо оставляется «как есть», либо заполняется исследователями на основании тех или иных соображений. В ряде случаев исследователь вставляет на место пропущенных данных какие-то осмысленные, на его взгляд, цифры и коды. В других случаях исследователи, стараясь избежать субъективного подхода, прибегают к различным методам восстановления пропущенной информации.

В работе были рассмотрены различные типы пропусков в данных и методы их восстановления. Для широкого применения рекомендован метод множественного восстановления данных, предложенный в конце 1970-х Дональдом Рубином. Метод достаточно сложен, как с точки зрения используемого математического аппарата, так и содержательной интерпретации результатов, что сдерживает его широкое применение в исследованиях, несмотря на его явные преимущества над другими методами восстановления данных. Например, он позволяет избежать систематических

¹ Ramaswamy V., Raghunathan T.E., Cohen S.H., Ozcan K. A Multiple Imputation Approach of Missing Data in Marketing Research: Unpublished working paper / University of Michigan, Department of Business Administration. 2001.

² Rubin D.B. Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations // Journal of Business & Economic Statistics. 1986. Vol. 4. No. 1.

ошибок в оценке коэффициентов регрессионных моделей, а также превосходит другие методы при большом проценте пропущенных данных. Удобная реализация данного метода в последних версиях широко распространенного статистического пакета SPSS позволяет рекомендовать его как наилучший на текущий момент метод восстановления данных.

Использование кластерного анализа в парадигме Data Mining для изучения структуры рынка труда

Мальцева Анна Васильевна
Алтайский государственный университет

Рынок труда является неизменно актуальной проблематикой научных изысканий различного дисциплинарного уровня. Эта тема представляется в целом безграничной ввиду постоянного развития объекта, что требует поиска все новых теоретических, методных и технологических решений анализа его состояний. Неизменной при любых трансформациях рынка труда, вызванных политическими или экономическими причинами, остается сущностная его характеристика — соотношение предложения и спроса рабочей силы. Такая постановка вопроса довольно очевидна и все же именно она может стать отправной точкой определения структуры рынка труда с несколько непривычной стороны: со стороны описывающих рынок труда данных, содержащихся в базах данных государственных служб занятости. Подобные хранилища эмпирического материала непривычны для социолога, но необходимы при анализе структуры рынка труда, поскольку содержат в себе всестороннее описание объекта исследования (исключением является разве что вероисповедание обращающихся граждан или наличия у них родственников за рубежом). *Благодаря такому подходу рынок труда как систему можно описать на макроуровне с помощью современных информационных технологий. Когда макроуровень будет взят в размерности, позволяющей выделять надындивидуальные структурные образования, масштаб которых ограничен системой рынка труда.*

Использование сведений из баз данных развивается в рамках информационного подхода к анализу данных, традиционно в бизнесе для описания бизнес-процессов¹. Данный подход позволяет изучать объект и содержащиеся в нем закономерности в его естественном состоянии. Это, кроме всего прочего, позволяет еще избежать директивности в определении базовой модели анализа объекта исследования, поскольку она в дальнейшем «подстраивается» под данные и используется для уточнения выявленных закономерностей. Информационный подход является одним из методологических оснований парадигмы анализа данных Data Mining или «глубокого анализа данных». В литературе данная парадигма еще называется методикой извлечения знаний, к которым также относят Knowledge Discovery in Databases (KDD, поиск знаний в базах данных). Если рассматривать методический вопрос с учетом последующих технологий реализации аналитических решений, то важно уточнить иерархию обозначенных выше терминов: «Data Mining является ядром методики» Knowledge Discovery in Databases². Процесс извлечения знаний из баз данных (KDD) — это процедура «получения из данных знаний в виде зависимостей, правил, моделей»³, включает в себя этапы выборки данных, очистки и трансформации, моделирования и интерпретации полученных результатов. При использовании парадигмы Data Mining возможно обнаружение «в сырых данных ранее неизвестных, нетривиальных»⁴ зависимостей или знаний. «Знания должны описывать новые связи между свойствами, предсказывать

¹ Колин К.К. Информационный подход как фундаментальный метод научного познания // Межотраслевая информационная служба / М.: ВИМИ, 1998. Вып. 1 (102). С. 3–17; Мальцева А.В. О результатах применения частной методики сегментации рынка труда // Вестник Самарского государственного университета. 2010. №7; Мальцева А.В. Социологический анализ рынка труда: поиск новых источников эмпирических данных // Сборник материалов VI Международной научно-практической конференции «Наука и современность – 2010» / Под общ. ред. С.С. Чернова. Новосибирск: Изд-во НГТУ, 2010. Ч. 2. С. 36–39; Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2010.

² Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2010. С. 40–43.

³ Паклин Н.Б., Орешков В.И., 2010. С. 41.

⁴ Паклин Н.Б., Орешков В.И., 2010. С. 42.

значения одних признаков на основании других»¹. В данной парадигме анализа данных возможно применение различных методов: классификации, регрессии, кластеризации, ассоциации² и ряд других. Реализация аналитического процесса в рамках данной парадигмы реализуется на основе аналитических платформ³. Данный класс программного обеспечения позволяет использовать наборы интегрированных инструментов анализа данных большой размерности и различных уровней измерения.

Осенью 2009 года авторским коллективом под научным руководством автора статьи был осуществлен аналитический проект для Управления по труду и занятости населения Администрации Алтайского края по сегментации официальной части открытого рынка труда⁴. Эмпирическим основанием проекта послужили сведения базы данных Управления, содержащие описание двух основных сторон рынка труда: предложения рабочей силы и спроса на нее. Эти сведения содержат официально зарегистрированные обращения граждан, ищущих работу и дополнительные услуги в сфере труда, а также работодателей, предлагающих вакансии или участвующих в государственных программах предоставления услуг гражданам в сфере трудоустройства и трудовой адаптации. Методическое решение данной задачи было построено с учетом использования метода кластерного анализа путем построения самоорганизующихся карт признаков (карты Кохонена). Технологическая реализация осуществлялась на базе аналитической платформы Deductor⁵ и выполнялась на основе автоматизированной обработки данных в сценарном режиме.

Метод многомерного анализа, куда входят кластерный анализ, распознавание образов и факторный анализ — наиболее действенный инструмент исследования социально-экономических процессов, описываемых большим числом характеристик. Кластерный анализ наиболее ярко отражает черты многомерного анализа в классификации, его неоспоримое преимущество в том, что он позволяет производить разбиение объектов не по одному параметру, а по целому набору признаков. Кроме того, кластерный анализ, в отличие от большинства математико-статистических методов не накладывает никаких ограничений на вид рассматриваемых объектов, позволяя рассматривать множество исходных данных практически произвольной природы. Это имеет большое значение, например, для сегментации разного рода рынков, когда показатели имеют разнотипный вид, затрудняющий применение традиционных эконометрических подходов⁶.

Сегментация рынка — деление (дифференциация) любого рынка на отдельные части (сегменты) с учетом множества критериев и факторов. Выявление структуры и скрытых закономерностей рынка часто может быть выполнено только на основе решения задачи кластеризации. Кластерный анализ — один из распространенных методов решения задач, возникающих при анализе социальной и экономической информации. Примером может служить описание структуры социально-территориальных групп, разбиение клиентов рынка на различные по уровню мотивации группы и т.д., т.е. нахождение групп схожих объектов в выборке данных. В результате решения задачи кластеризации выявляется скрытая структура данных и обнаруживаются неочевидные закономерности в данных. Кластерный анализ — это способ группировки многомерных объектов, основанный на представлении результатов отдельных наблюдений точками подходящего геометрического пространства с последующим выделением групп как «сгустков» этих точек (кластеров, таксонов)⁷. «Кластер» (cluster) в английском языке означает «сгу-

¹ Паклин Н.Б., Орешков В.И., 2010. С. 42.

² Berry M.J.A., Linoff G.S. Data Mining techniques. Ind., Indiana: Wiley, 2004.

³ Мальцева А.В. Современные информационные технологии в обеспечении прикладных исследовательских задач в социально-культурной сфере // Известия Алтайского государственного университета, сер. «Математика и механика. Управление, вычислительная техника и информатика. Физика». 2009. №1(61). С. 143.

⁴ Мальцева А.В., Чудова О.В., Шилкина Н.Е. Сегментация рынка труда: теория и методика. Барнаул: АзБука, 2010; Мальцева А.В. Социологический анализ рынка труда: классические теории и новейшие технологии. Барнаул: Изд-во Алт.ГУ. 2010; Мальцева А.В. Использование современных методов и технологий анализа данных для принятия управленческих решений (на примере решения задачи сегментации рынка труда) // Вестник Евразийской академии административных наук. 2010. №2.

⁵ www.basegroupplabs.ru

⁶ Паклин Н., Орешков В. Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2009. С. 624.

⁷ Кузнецов Д.Ю., Трошина Т.Л. Кластерный анализ и его применение // Ярославский педагогический вестник. 2006. Вып. 4. С. 103–107.

сток», «гроздь винограда», «скопление звезд» и т.д. Данный метод исследования получил развитие в последние годы в связи с возможностью компьютерной обработки больших баз данных. Многообразие алгоритмов кластерного анализа обусловлено множеством различных критериев, отражающих те или иные аспекты качества автоматической группировки. При этом практически все методы характеризуются высокой вычислительной сложностью и чувствительностью к точности задания априорной информации о количестве кластеров или максимально допустимом размере кластера, о предполагаемой структуре кластеров¹.

Предполагаемая выборка объектов исследования $O_1, \dots, O_n$ (участники рынка труда: вакансии или клиенты службы занятости) может быть сформирована в результате отбора некоторых представителей генеральной совокупности. Использование методов Data Mining осуществляется для выборки данных, содержащих n записей. Из нее формируется k классов (групп объектов); число классов может быть, как выбрано заранее, так и не задано (в последнем случае оптимальное количество кластеров должно быть определено автоматически). Решение задачи кластеризации предполагает на основании анализа параметров объекта $P_1, \dots, P_m$ (данных об участниках рынка из баз данных) выделение схожих объектов и представление результата в форме, удобной для восприятия. Набор $P_1, \dots, P_m$ может включать переменные разных типов (количественные, качественные, порядковые). Решением задачи кластерного анализа является разбиение, удовлетворяющее некоторому условию оптимальности. Этот критерий может представлять собой некоторый функционал, выражающий уровни желательности различных разбиений и группировок.² Этот функционал часто называют целевой функцией. Задачей кластерного анализа является задача оптимизации, т.е. нахождение минимума целевой функции при некотором заданном наборе ограничений³.

Кластерный анализ включает в себя следующие этапы⁴: 1) получение репрезентативной выборки, формирование системы переменных и выделение наиболее информативных признаков; 2) вычисление некоторой меры близости между объектами (признаками). Это означает переход от таблицы X «объект-признак» к таблице «объект-объект» $S = \|s_{ij}\|_{n \times n}$ (таблице «признак-признак» $R = \|r_{ij}\|_{k \times k}$), где s_{ij} – мера близости между объектами O_i и O_j (r_{ij} – мера близости между объектами P_i и P_j); 3) применение конкретного метода кластерного анализа; 4) проверка достоверности и интерпретация полученных результатов. Каждый из перечисленных этапов играет важную роль при использовании кластерного анализа для решения прикладных задач. Выделяется несколько подходов к решению задачи кластерного анализа, которые основаны на различных представлениях о задаче, использовании специфичной для каждой предметной области дополнительной информации и т.д. Среди них вероятностный подход⁵; подход, использующий аналогию с центром тяжести⁶; подход, основанный на понятии ближайшего соседа⁷; иерархический подход; нечеткие алгоритмы кластерного анализа⁸; подход, использующий ис-

¹ Чудова О.В. Кластерный анализ социологической информации // Материалы IV всероссийской научной конференции «Сорокинские чтения» «Отечественная социология: обретение будущего через прошлое». М., 2008. С. 89–92.

² Кузнецов Д.Ю., Трошина Т.Л. Кластерный анализ и его применение // Ярославский педагогический вестник. 2006. Вып. 4. С. 103–107.

³ См.: Мальцева А.В., Чудова О.В., Шилкина Н.Е. Сегментация рынка труда: теория и методика. Барнаул: Азбука, 2010

⁴ Бритвихин В.А., Красина Ф.А., Симонцев С.Н. Использование кластерного анализа для типологизации признаков // Социологические исследования. 1994. № 2. С. 100–105.

⁵ Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика, 1989. С. 607.

⁶ Дуда Р., Харт П. Распознавание образов и анализ сцен. М.: Мир, 1976. С. 512.

⁷ Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика, 1989. С. 607.

⁸ Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах. Киев: Слово, 2008. С. 344.

кусственные нейронные сети¹.

В рамках решаемой задачи наиболее полезно применить методы позволяющие проводить разделение на не пересекающиеся классы с произвольным их числом. Кластеризация методом Кохонена в данном случае имеет ряд преимуществ: возможно решение задач большой размерности, число кластеров может быть не задано, формируют наглядное двумерное отображение множества объектов и т.д. Метод анализа с использованием самоорганизующихся карт признаков (СКП) Кохонена², являющихся разновидностью неуправляемых нейросетей, позволяет автоматизировать все действия по поиску закономерностей. Главной причиной использования СКП для исследовательского анализа данных и извлечения знаний является то, что данный метод не требует никаких априорных предположений о распределении данных; позволяет обнаруживать в наборах данных неизвестные ранее структуры или образы при помощи обучения без учителя. Он был предложен Тьюво Кохоненом в начале 1980-х гг. и нашел широкое применение в инженерной области (для распознавания речи, в робототехнике и др.). Технология СКП представляет собой набор аналитических процедур и алгоритмов, позволяющих преобразовать традиционное описание множества объектов, заданных в многомерном ($n > 3$) пространстве признаков плоской базы данных, в двумерную карту, устроенную таким образом, что близким объектам в многомерном пространстве отвечают рядом стоящие точки (их образы) на карте³. В результате трудно анализируемые в совокупности многомерные объекты получают простой и наглядный вид на двумерной карте, которая сохраняет их основные свойства (топологию и распределение в многомерном пространстве). Применение технологии СКП дает ряд преимуществ: обнаружение групп объектов с одинаковыми характеристиками (кластерами) по их локализованному расположению на специально создаваемой карте кластеров; проверка содержательного описания обнаруженных групп по специфическим особенностям, обнаруженным на карте признаков, а также на проекциях карты кластеров на каждый признак в отдельности; выявление неявных связей и закономерностей между признаками; проведение оценки объектов в динамике, оценка изменений как в целом по структуре кластеров, так и по отдельности; позиционирование на карту новых объектов для придания им статуса (рейтинга); прогнозирование значений одних признаков объектов через другие; фильтрация объектов за счет поисковых уникальных критериев, формируемых в терминах СКП.

Карты Кохонена позволяют также представить полученную информацию в простой и наглядной форме путем нанесения раскраски⁴. Карта Кохонена позволяет визуализировать результаты кластеризации, в том числе и многомерные. Карта Кохонена состоит из сегментов прямоугольной или шестиугольной формы, называемых ячейками. Каждая ячейка связана с определенным выходным нейроном и представляет собой «сферу влияния» данного нейрона. Распределение векторов весов нейронов карты получается так же, как и в обычной сети Кохонена, т.е. на основе конкурирующего обучения. Объекты, векторы признаков которых оказываются ближе к вектору весов данного нейрона, попадают в ячейку, связанную с ним. Тогда распределение объектов на карте в целом соответствует распределению векторов весов нейронов. Следовательно, если объекты на карте расположены близко друг к другу, то и векторы признаков этих объектов близки и наоборот, если ячейки с объектами находятся далеко друг от друга, то и векторы их признаков различаются сильно. Хотя расстояние между объектами уже позволяет сделать выводы о степени их сходства или различия, но важна информация о том, в чем проявляется это сходство и различие, по каким признакам они различаются в наибольшей степени, а по каким — в наименьшей и т.д. Таким образом, есть два важных фактора: положение объекта на карте (расстояние до других объектов) и цвет ячейки. Очевидно, что при таком способе ви-

¹ Круглов В.В., Дли М.И., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. М.: Физматлит, 2000. С. 224; Оссовский С. Нейронные сети для обработки информации / Пер. с польского И.Д. Рудинского. М.: Финансы и статистика, 2002. С. 365.

² Дебок Г., Кохонен Т. Анализ финансовых данных с помощью самоорганизующихся карт. М.: Альпина Паблишер, 2001. С. 317.

³ Горбунов А. Применение самоорганизующихся карт в бизнесе и финансах // Банковские технологии. 1999. № 4. URL: [http://it2b2.view2.page103.html]

⁴ См.: Паклин Н., Орешков В. Бизнес-аналитика: от данных к знаниям. СПб.: Питер, 2009.

зуализации на одной карте можно использовать расцветку только по одному признаку, т.е. для значений нескольких признаков, следует строить отдельные карты по каждому из признаков. Например, если объекты имеют признаки «Возраст» и «Пол», то для получения полной информации о значениях этих признаков необходимо построить две карты, отражающие соответствующие признаки¹.

Наиболее востребованным и популярным сегодня направлением реализации аналитических методов кластерного анализа являются нейронные сети. Их популярность объясняется тем, что они более эффективно решают те задачи, которые бессильны решить традиционные линейные статистические модели. Нейронные же сети в своей основе нелинейны, что позволяет учесть нелинейные, более глубокие связи между исходными данными и результатами. Под нейронными сетями (НС) подразумеваются вычислительные структуры, которые моделируют простые биологические процессы, обычно ассоциируемые с процессами человеческого мозга. Адаптируемые и обучаемые, они представляют собой распараллеленные системы, способные к обучению путем анализа положительных и отрицательных воздействий. Элементарным преобразователем в данных сетях является искусственный нейрон или просто нейрон, названный так по аналогии с биологическим прототипом. Большинство моделей НС требует обучения, то есть подбора таких параметров сети, при которых она лучше всего справляется с поставленной проблемой.

Метод кластерного анализа с использованием самоорганизующихся карт Кохонена является разновидностью неуправляемых нейросетей, позволяющих автоматизировать все действия по поиску закономерностей, что особенно важно для совокупностей данных большой размерности. Самоорганизующиеся карты признаков (СКП) представляют собой не только эффективный алгоритм кластеризации, но и позволяют представлять ее результаты в виде двухмерных карт, где расстояния между объектами соответствуют расстоянию между их векторами в многомерном пространстве, а сами значения признаков отображаются различными цветовыми оттенками.

Стандартной аналогией является сравнение с обычной географической картой, где размещение объектов и расстояния между ними соответствуют их расположению на земной поверхности. В картах кроме горизонтальных координат, необходимо показать и рельеф, высоту гор, холмов, а также глубину водоемов. Для этого используется специальная гамма. Так высота местности отображается с помощью оттенков коричневого, глубина морей и океанов — синего, то есть чем выше объект, тем более темным цветом он окрашивается. Таким образом, двумерная карта позволяет показывать три измерения. Карта Кохонена состоит из сегментов прямоугольной или шестиугольной формы (для удобства пользователя), называемых ячейками. Каждая ячейка связана с определенным выходным нейроном и представляет собой «сферу влияния» данного нейрона. Распределение векторов весов нейронов карты получается на основе конкурирующего обучения. Объекты, векторы признаков которых оказываются ближе к вектору весов данного нейрона, попадают в ячейку, связанную с ним. Тогда распределение объектов на карте в целом соответствует распределению векторов весов нейронов. Следовательно, если объекты на карте расположены близко друг к другу, то и векторы признаков этих объектов близки и наоборот, если ячейки с объектами находятся далеко друг от друга, то и векторы их признаков различаются сильно. Хотя расстояние между объектами уже позволяет сделать выводы о степени их сходства или различия, но важна информация о том, в чем проявляется это сходство и различие, по каким признакам они различаются в наибольшей степени, а по каким — в наименьшей. Специальная раскраска карт и позволяет ответить на эти вопросы, выполняя функции третьего измерения. Идея состоит в том, что каждой ячейке на карте назначается цвет в соответствие со значениями признаков объектов в них. Таким образом, есть два важных фактора: положение объекта на карте (расстояние до других объектов) и цвет ячейки. Данный способ визуализации предполагает подробное рассмотрение в рамках одной карты только одного признака².

¹ Мальцева А.В. Сегментация рынка труда с использованием интеллектуальных алгоритмов анализа данных // Материалы XVII международной конференции студентов, аспирантов и молодых ученых «Ломоносов–2010» / Отв. ред. И.А. Алешковский, П.Н. Костылев. М.: МГУ; Мысль, 2010.

² См.: Мальцева А.В., Чудова О.В., Шилкина Н.Е. Сегментация рынка труда: теория и методика. Барнаул: Азбука, 2010.

Важно отметить, что сегментация целевой аудитории методом кластерного анализа с применением карт Кохонена используется преимущественно как инструмент стратегического анализа и уточнение статистических показателей по отдельным картам и/или сегментам (кластерам) не будет достаточным. Результат позволит увидеть картину в целом и является идеальной отправной точкой для дальнейшего подробного анализа данных. В частности, для понимания правил формирования тех или иных сегментов необходимо применять другие методы анализа, например, построение деревьев решений. Для решения задачи классификации с неопределенными начальными данными на практике наибольшее распространение получил метод дерева решений, который позволяет определить набор правил классификации типа «Если — То». Метод дерева решений — это один из методов автоматического анализа огромных массивов данных¹.

В основе данной технологии лежит идея последовательного разбиения множества объектов на подмножества таким образом, чтобы значения зависимой переменной в каждом из них были как можно более однородными. На каждом шаге разбиение производится только по одной независимой переменной, которая делает его наилучшим. По завершении всего процесса получается дерево последовательных разбиений, по существу — дерево решений, то есть набор правил выбора в иерархической структуре. Каждый путь от вершины до листа дерева (конечного узла) образует правило².

На сегодняшний день существует значительное количество алгоритмов, которые реализовывают деревья решений CART, C4.5, NewId, ITrule, CHAID CN2 и т.д. В данной работе для разработки модели принятия решений будет использоваться алгоритм C4.5, так как он позволяет работать с числовыми, категориальными, пропущенными и зашумленными данными³. Рассмотрим основные положения алгоритма C4.5. Пусть имеется множество примеров $\hat{O}$, разделенных по значениям классифицирующего признака на полные непересекающиеся классы $C_1, C_2 \dots C_k$, где каждый элемент этого множества описывается m атрибутами. Количество примеров в множестве $\hat{O}$ будем называть мощностью этого множества и будем обозначать $|\hat{O}|$. Задача заключается в построении иерархической классификационной модели в виде дерева из множества примеров $\hat{O}$. Процесс построения дерева происходит сверху вниз. Сначала создается корень дерева, затем потомки корня и т.д. На первом шаге мы имеем пустое дерево (имеется только корень) и исходное множество $\hat{O}$ (ассоциированное с корнем). Требуется разбить исходное множество на подмножества. Это можно сделать, выбрав один из атрибутов в качестве проверки⁴.

Примененный комплекс методик многомерного анализа логико-алгебро-геометрического направления позволяет устанавливать корреляции, тенденции, взаимосвязи и закономерности на рынке труда посредством математических и статистических алгоритмов кластеризации, создания субвыборок, регрессионного и корреляционного анализа. Разработанный количественный инструмент является наиболее информативным в исследованиях социально-экономических процессов, описываемых большим числом характеристик. Использование парадигмы Data Mining дало возможность представить данные в виде, четко отражающем процессы, происходящие на рынке труда, а также построить модель, при помощи которой можно анализировать кризисные ситуации. Применение данных технологий позволило получить принципиально новые возможности для оценки рынка труда, для выявления значимых классов профессий, предлагаемых и требующихся на рынке, выявления факторов, влияющих на динамику рынка труда в Алтайском крае, автоматизации этапов, необходимых для осуществления сегментирования рынка труда с высокой точностью и скоростью. Внедрение данных разработок решает проблему выстраивания единой системы работы с информацией, в ежедневно обновляющейся базе данных.

¹ Берестнева О.Г., Муратова Е.А. Построение логических моделей с использованием деревьев решений // Известия Томского политехнического университета. 2004. Т. 207. Вып. 2. С. 55–61.

² См.: Кацко И.А., Паклин Н.Б. Практикум по анализу данных на компьютере. М.: Колосс, 2009.

³ Деревья решений – C4.5 математический аппарат // Basegroup Labs. Технологии анализа данных. [URL: <http://www.basegroup.ru/library/analysis/tree/>]

⁴ См. Мальцева А.В., Чудова О.Н., Шилкина Н.Е., 2010.

Информационные технологии анализа общественного мнения и маркетинга

Мхитарян Владимир Сергеевич, *НИУ ВШЭ*

Черепанов Евгений Васильевич,

Институт экономики и комплексных проблем связи

Специфика использования вероятностно-статистического формализма в социологических исследованиях связана с тем, что обычно в них имеют дело с анализом качественных признаков (нечисловых показателей) объектов изучения¹. При этом каждый объект может обладать целым множеством таких признаков, т.е. правомерно говорить о *многомерных наблюдениях* (элементах заданного множества людей).

Для анализа общественного мнения используются методики опроса населения, теоретически основанные на выборочном методе. По своей сути выборочный метод сводится к количественному измерению заданного свойства на некоторой части общего («выборке» из изучаемой генеральной совокупности). Научное обоснование выборочного метода основано на асимптотических свойствах выборочных статистических оценок, которые правомерны только для абсолютно однородных генеральных совокупностей².

Ясно, что для переноса вывода о свойствах частного (выборки) на свойства общего (генеральной совокупности) требуется, чтобы это частное в широком смысле хорошо «представляло» общее. Практически это означает, что структура выборки в многомерном плане должна быть хорошо согласована со структурой генеральной совокупности. А добиться этого на практике очень непросто. Суть дела в том, что любой социум является заведомо неоднородным, причем по многим априорным классификациям, множеством. В этой ситуации у исследователей остается только два пути: строить квотные выборочные ансамбли, приближенно похожие по своей структуре на генеральную совокупность, или математически строго учесть различие структур выборки и генеральной совокупности.

Традиционные технологии 1-го поколения (работа с квотными выборками)

Весь социологический мир работает с *квотными выборками*. Суть этой методики сводится к искусственному конструированию такой выборки, которая бы по нескольким базовым признакам (иногда называемым «социологическим паспортом») достаточно точно отражала структуру генеральной совокупности. При этом на практике социологических и маркетинговых работ исследователь сталкивается со следующими принципиальными трудностями:

- при больших мощностях генеральной совокупности и работе на больших территориях (региональные и российские опросы) очень трудно практически обеспечить действительно квотную выборку уже по трем признакам социологического паспорта;
- разовое формирование квотной выборки — чрезвычайно дорогостоящее и длительное мероприятие. Но и поддержание постоянной и периодически используемой (однажды сформированной) квотной выборки является дорогим удовольствием. Причем такая выборка не обеспечивает объективность опросов и связана с неконтролируемыми процессами изменения ее состава (болезни, командировки) и потенциальной неискренностью постоянных респондентов, получающих за работу постоянную материальную компенсацию;
- при работе с квотными выборками всегда остается «открытым» вопрос оценки погрешности получаемых результатов. Все публикуемые в СМИ «погрешности в 3 %» — большая натяжка. Для их вычисления используются вероятностные соотношения, базирующиеся на поня-

¹ Мхитарян В.С., Черепанов Е.В. Проблемы прикладной социологии в их привязке к социально-экономическим исследованиям // Информатика, социология, экономика, менеджмент: Межвузовский сборник научных трудов. М.: Академия менеджмента инноваций, 2006. Вып. 3. Ч. 2. С. 23–33; Черепанов Е.В. Вероятностно-статистические основы прикладной социологии и маркетинговых исследований. М.: Академия менеджмента инноваций, 2006.

² Мхитарян В.С., Черепанов Е.В. Стохастические методы в прикладных исследованиях: корректность и надежность использования // Анализ социально - экономических и политических процессов и систем. М.: Академия менеджмента инноваций, 2006. Вып. 3. С. 57–67; Черепанов Е.В. К вопросу корректности использования стохастического формализма в социологических и социально-экономических исследованиях // Безопасность Евразии. 2007. № 2 (28). С. 386–402.

тии случайности событий и величин, что при работе с квотной (т. е. заведомо не вполне случайной) выборкой требует строгого обоснования¹.

• И, наконец, серьезно говорить об оперативности быстрых замеров общественного мнения (экспресс-опросов), основанных на работе с квотными выборками, при изучении большой численности населения на больших территориях, просто невозможно.

Технологии 2-го поколения (работа со случайными выборками)

Альтернатива квотному отбору при изучении общественного мнения — случайный отбор². Этот путь был предложен в конце 1980-х годов прошлого века. Подход основан на оригинальном методе вычисления условных вероятностей, который, в свою очередь, базируется на аппарате исчисления статистик бинарных отношений на множествах³. В течение 1990–2000-х годов аппарат был развит до уровня комплексной методологии⁴ и апробирован при выполнении около 30 реальных проектов электорального и маркетингового характера. Соответствующая информационная технология имеет строгое вероятностно-статистическое обоснование.

В основе методики лежит компьютерный анализ структуры случайной выборки и ее сопоставление со структурой соответствующей генеральной совокупности. При этом вычислительный процесс организован так, что обеспечивает репрезентативность результатов на этапе расчетов. Математически строго показано, что учет структуры заданной генеральной совокупности и ее формальное сопоставление со структурой выборки сводятся к исчислению условных вероятностей для распределений гипергеометрического типа. Их значения позволяют обеспечить высокую точностью оценок.

В многомерном формально-математическом смысле исчисление соответствующих условных вероятностей обеспечивает репрезентативность результатов по любому числу признаков в социологическом паспорте. Более того, на основе теории неравноточных измерений, строго показано, что при данной технологии итоговая оценка имеет тем меньшую погрешность, чем большее число признаков включено в социологический паспорт.

Технология работы со случайными выборками позволяет:

- резко повысить точность получаемых результатов;
- существенно повысить оперативность проведения исследования;
- радикально повысить полноту описания, дав различные «срезы» изучаемой ситуации;
- значительно снизить и оптимизировать финансовые затраты на исследования.

Практически, при объемах выборки порядка 1 500 – 2 000 респондентов, методика обеспечивает погрешности:

- порядка 0,5–1,0 % для населения в целом;
- от 2,5 до 3,5 % для социально-демографических категорий населения.

Технологии 3-го поколения (групповое анкетирование на «малых выборках»)

Данная технология предназначена для оценки мнения населения по различным структурно и логически сложным проблемам (тематикам). Такими проблемами являются тематические информационные блоки, отношение к которым невозможно достоверно замерить с помощью массового социологического опроса населения:

¹ Подробнее см.: Черепанов Е.В. Стохастическое описание выборочного метода // Социология: методы, методология, математическое моделирование. 2007. № 25. С. 167–189.

² Мхитарян В.С., Черепанов Е.В. Выборочный метод на случайных выборках в социологических и социально-экономических исследованиях: стохастическое обоснование // Информатика, социология, экономика, менеджмент: Межвузовский сборник научных трудов. М.: Академия менеджмента инноваций, 2007. Вып. 4. Ч. 2. С. 38–47; Мхитарян В.С., Черепанов Е.В. Выборочный метод на случайных выборках в социологических и социально-экономических исследованиях: статистические оценки // Информатика, социология, экономика, менеджмент: Межвузовский сборник научных трудов. М.: Академия менеджмента инноваций, 2007. Вып. 5. Ч. 2. С. 42–61.

³ Черепанов Е.В. и др. Статистики бинарного отношения на множествах // Проблемы перспективного планирования и управления: Сборник научных трудов. М.: Изд. Госплана СССР, 1990. С. 88–98.

⁴ Черепанов Е.В. Статистическая методология для задач социологических и социально-экономических исследований. М.: Академия менеджмента инноваций, 2007; Черепанов Е.В. Стохастические методы прикладной социологии и маркетинга рынков. М.: Академия менеджмента инноваций, 2008.

- вследствие трудной воспринимаемости проблемы, для понимания которой требуются подробные разъяснения, образные демонстрации;
- вследствие малой известности широким слоям населения подробностей тематики;
- из-за отсутствия у большинства населения выраженного отношения к изучаемой теме;
- вследствие необходимости специального моделирования реальных ситуаций для выявления возможного отношения населения (после проведения информационных и рекламных мероприятий).

Технология «малых выборок»

В основу технологии ⁴ положена авторская методика, основанная на выявлении количественной меры ¹ связей между известными элементами таблиц и восстановлении пропущенных элементов таблиц по выявленным взаимосвязям. Эта проблематика в прошлом серьезно изучалась рядом специалистов, но в опубликованных социальных исследованиях, к сожалению, доныне не нашла сколько-либо серьезного внедрения. Разработанный нами подход основан на двух типах данных, полученных в результате опросов населения.

Первый тип данных — результат социологического опроса населения по вопросам, отражающим общие, понятные неподготовленному человеку, стороны исследуемой проблематики. Опрашивается не менее 1,5–2,0 тыс. человек, представляющих все основные социально-демографические группы населения. В данной части работы исследуются общие (качественные) положения, в которых отражается изучаемая тематика.

Второй тип данных — результат опроса нескольких малых групп населения (по 30–60 человек каждая) по вопросам подробной анкеты. В ней содержатся вопросы дополняющие, детализирующие и расшифровывающие общие (качественные) положения, исследуемые «большим» социологическим опросом.

В процессе беседы участникам малой группы объясняется смысл «тонких» вопросов, проводятся необходимые разъяснения, доводится новая информация. Здесь возможно выявить причины тех или иных ответов на вопросы социологического опроса, оценить влияние новой информации на формирование (изменение) отношения к проблеме.

Оба выборочных ансамбля участников опросов формируются случайным образом. Репрезентативность по социально-демографическим группам обеспечивается на этапе обработки двух массивов анкет. Для этого используются числовые данные Госкомстата о структуре населения изучаемого региона.

При расчетах ^{4, 2} на первом этапе результаты опроса малой группы распространяются на участников социологического опроса населения. Учитываются и статистические взаимосвязи участников опросов, и взаимосвязи вопросов, задаваемых при этих опросах.

Используемый метод совмещения двух блоков данных имеет строгое вероятностно-статистическое обоснование. В результате мы получаем детальное мнение участников малой группы по вопросам, заданным в процессе беседы участникам группового анкетирования.

Вторым этапом расчета является «пересчет» таких оценок на все население России (или заданного изучаемого региона) и его основные социально-демографические категории. Подчеркнем, что речь идет именно об аппроксимации результатов на население и его категории, а не о каких-то «процентах от опрашиваемых».

Построение «малых выборок»

Несколько упрощенно этапы работы выглядят следующим образом.

1-й этап (опрос населения). По интересующей нас тематике на простом (доступном неподготовленному человеку) уровне проводится опрос населения. Технически он проводится по технологии, описанной выше, и охватывает 1,5–2,0 тыс. человек, которые отбираются случайным

¹ Черепанов Е.В. О функции сходства и типологическом пространстве в маркетинговых и социальных исследованиях // Материалы IV конференции «Современные проблемы формирования методного арсенала социолога. М.: ГУ–ВШЭ; РОС, 2010.

² Азаров С.В., Черепанов Е.В. Статистические оценки при групповом анкетировании методом «малых выборок» // Информатика, социология, экономика, менеджмент: Межвузовский сборник научных трудов. М.: Академия менеджмента инноваций, 2006. № 3. С. 38–46.

образом. В анкете задействуются простые вопросы по интересующей нас тематике, которые не требуют от неподготовленного среднего человека дополнительных знаний. В итоге мы не получаем глубокой картины общественного мнения, а лишь «снимаем» только некие общие настроения по интересующей нас тематике. По сути, мы получаем некоторый расширенный социологический паспорт для последующей работы с «малыми выборками» респондентов.

2-й этап (работа с малыми выборками населения). В зале собирается небольшое (30–50 человек) количество респондентов, представляющих интересующее нас население. Работа с ними напоминает нечто среднее между традиционными «фокус-группами» и обычным школьным уроком. Вначале (до «урока») малой выборке респондентов предлагают заполнить уже знакомую нам с предыдущего этапа простую тематическую анкету. «Урок» разделен на логически законченные части. Изложив слушателям заданную версию данной части мероприятия, социолог, исполняющий роль «учителя», задает респондентам контрольные вопросы. Затем «учеников» просят заполнить соответствующую часть «тонкой» анкеты. В итоге, за 3–4 часа вся «тонкая» анкета заполнена каждым из участников малой выборки.

3-й этап (оценка элементов матрицы «большого опроса»). По статистическим связям между столбцами и строками матрицы данных можно оценивать характеристики объектов самых разных природы и назначения. Это обусловлено тем, что соответствующие алгоритмы полностью универсальны.

4-й этап (аппроксимация результатов на генеральную совокупность). Проводится по методике и технологии, описанным в п. 2. В результате получается искомая количественная оценка для всего изучаемого населения. Фактически мы получили оценку конечного результата агитационных мероприятий при заданной тактике (например, политической или рекламной кампании). И задали мы эту тактику сами. Испробовав на малых выборках несколько вариантов точек зрения до населения, мы можем отобрать лучшие, наиболее эффективные из аспектов и составляющих каждой тактики.

5-й этап (выработка рекомендаций и планов). Проводится на основе экспертно-аналитической обработки и обобщения итоговых результатов проведенной работы.

Процедуры сбора данных по направлению «Retail and Technology» в компании «GfK»

Ревякин Сергей Анатольевич,
Международный институт маркетинговых
и социальных исследований «ГфК-Русь»

Понятие «панель» было заимствовано из американского английского¹. Дальнейшая этимологическая цепочка теряется. Возможно, оно произошло от одноименного голландского слова, которое, среди прочего, означает возделывание земли. Понятие также существует в американской официальной и парламентской терминологии. Оно также может означать документ, содержащий имена членов комиссии, или, напротив, список имен членов конгресса, являющихся кандидатами для назначения в комитет конгресса.

На сегодняшний день панельные исследования являются одним из наиболее важных источников данных при проведении маркетинговых исследований. В таблице 1 указана их классификация в зависимости от используемой методологии².

Одним из видов панельных исследований является ритейл-аудит. Ритейл-аудит (retail audit), аудит розничной торговли — это вид маркетинговых исследований, который предполагает мониторинг таких параметров розничной торговли, как цены, ассортимент, дистрибуция марок в торговых точках, объемы продаж по номенклатуре и другие параметры, определяемые требованиями заказчика или особенностями исследуемого рынка. Впервые метод был апробирован в 1923 году Артуром Ч. Нильсоном³.

¹ GfK Panelguide Retail and Technology, 7th Edition. Management/Publisher: G. Redwitz. P. 159.

² GfK Panelguide Retail and Technology, 7th Edition. Management/Publisher: G. Redwitz, p.160

³ URL: <http://www.oirom.ru/en/events.php?id=12> (Дата обращения - 25.01.2010)

Таблица 1

Классификация методологий маркетинговых исследований

Критерий	Характеристики	
Тип сбора данных	Первичный (полевое исследование)	Вторичный (кабинетное исследование)
Временной период	Повторяющиеся	Межсекционный анализ
Метод сбора данных	Наблюдение	Интервью Эксперимент
Масштаб исследования	Выборка	Генеральная совокупность
Предмет исследования	Статистика по рынку	Демография

Источник: GfK Panelguide Retail and Technology, 7th Edition

Ритейл-аудит условно можно разделить на внешний и внутренний. К внешнему ритейл-аудиту можно отнести сторчек (store check), который представляет собой регистрацию (обследование) ассортимента товара, цен, рекламной активности производителей, размещения товара на полках и прочих внешних атрибутов продаж, которые можно зафиксировать методом наблюдения. При этом для сбора данных используются такие методы, как интервью, телефонный опрос, наблюдение, анкетирование в точках продаж и др.

Сторчек как метод обычно направлен на проверку эффективности работы мест продаж. Он проверяет, насколько выбранная маркетинговая стратегия компании реализуется на местах продаж, в том числе промоушен и сервис, а также позволяет сравнить маркетинговые трейд-стратегии. И чем больше сеть магазинов, тем больше могут быть отличия между ними, и тем больше отклонения от выбранной маркетинговой стратегии.

При этом непосредственные исполнители в магазинах зачастую в силу ряда причин не могут произвести качественный внешний срез происходящего в магазине. В связи с этим появляется необходимость внешнего аудита, который позволит более объективно понять, насколько реализуется маркетинговая стратегия торговой компании в конкретной точке продаж.

Сторчек начинается с согласования между заказчиком и исполнителем товарных позиций, по которым будет проводиться проверка. При этом, как правило, предмет сторчека — параметры продаж, которые являются ключевыми факторами успеха для данной торговой марки и обеспечивают успех на рынке. К ним относятся ассортимент продукции, оформление мест продаж (внешнее и внутреннее), качество обслуживания и др.

К внутреннему ритейл-аудиту можно отнести прямое получение данных о продажах от ритейлеров и предоставление их производителям на основании подписки (сторчек также бывает как одноразовым, так и периодическим). Этот вид ритейл-аудита является наиболее эффективным с точки зрения достоверности получаемой информации. Он используется компаниями, специализирующимися на исследовании тенденций (динамики) продаж розничного рынка, таких как «ACNielsen», «Synovate» и «GfK».

Притом если сторчек по типу использования, обычно, одноклиентный, то внутренний ритейл-аудит, как правило, носит синдикативный характер.

Таким образом, обычно внутренний ритейл-аудит больше ассоциируется с созданием сплошных панельных данных по магазинам методом прямого сбора данных «из кассы», а сторчек — с созданием дискретного массива данных по текущей рыночной ситуации, и по методу сбора больше схож с «mystery shopping».

Стандартный процесс внутреннего ритейл-аудита в компании «GfK» может описать следующим алгоритмом (рис. 1)¹:

¹ Компания «ГФК Русь» — дочернее предприятие всемирно известного маркетингового института GfK Group, который работает в более чем 100 странах мира (крупнейший в Европе и один из крупнейших в мире). При этом покрытие по исследуемым рынкам достигает 85–90%, и в выборке имеется более 16 тыс. магазинов разного профиля по всему миру. Всего в «GfK Group» мониторится более 300 продуктовых групп.

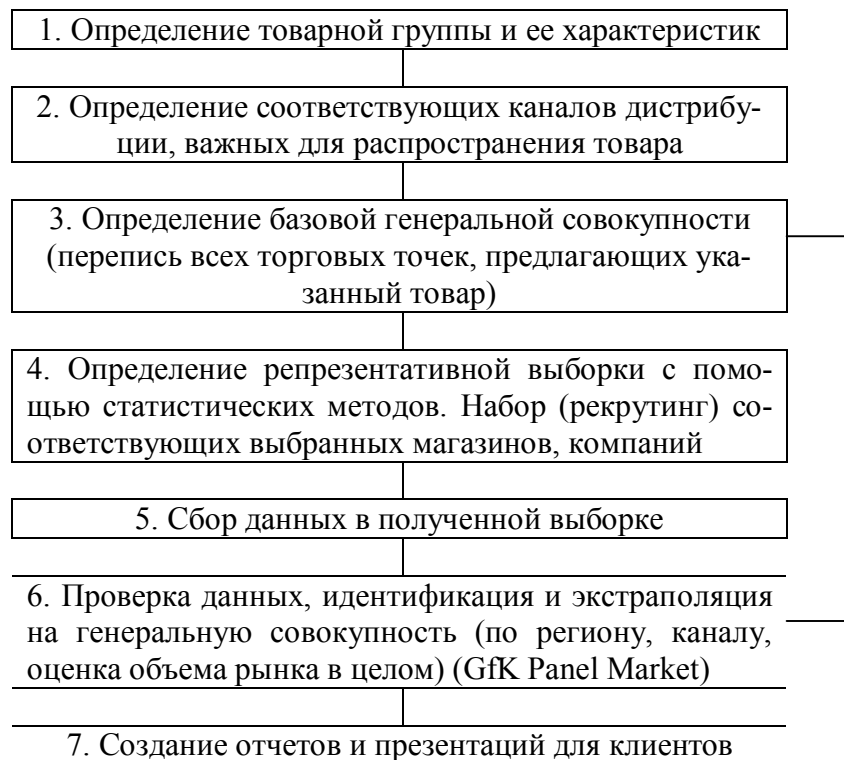


Рис. 1. Стандартный процесс создания ритейл-панели.

Такой алгоритм применим к любым товарам: печенье и компьютерам, молоку и утюгам, сыру и холодильникам. Главное — определить, что и как детально мониторить.

Классический внутренний ритейл-аудит — это мониторинг, а следовательно и сбор данных, по каждому артикулу (как иногда говорят специалисты, SKU — Stock Keeping Unit, или по-русски, «единица продаж»). Понятно, что в случае с молоком это будут упаковки в виде пакетов, бутылок и пр. Важно также при сборе данных не упустить, какого объема данный SKU. Это необходимо при дальнейшем пересчете в объемы проданного товара в литрах (как в случае с молоком). В случае с техникой или иными штучными товарами все проще: единица, она и есть единица, иначе «модель».

После того, как продуктовая группа, которая будет мониториться, определена в терминах продуктовых характеристик, описана должным образом, и описание согласовано с клиентом (клиентами), определяются релевантные каналы дистрибуции, в которых данные товары физически продаются. Определяется общее количество магазинов (структура, число, региональное распределение) в этих каналах. На основе этой информации формируется репрезентативная выборка, и рекрутируются соответствующие магазины, используя метод квотирования.

Одним из важных подготовительных этапов является определение частоты сбора данных. Необходимо собирать данные о продажах в определенные сроки, на постоянной основе. Иногда это может быть двухмесячный период, если сезонность диктует такую необходимость. Пример — солнцезащитные очки; по ним не так много «пиков» продаж. Такая же частота установлена и по рынку автомобильных шин в России.

В то же время в постоянно меняющемся и динамичном розничном «мире» один из ключевых факторов развития — возможность иметь самые свежие сведения о состоянии и движении рынка, реакции на изменения и измерения эффекта реакции. В Японии и Франции по рынку фильмов, музыки и игр сбор данных осуществляется ежедневно в связи с особенностью рынка (игрокам важна динамика продаж с момента выпуска в прокат).

Наконец, наступает этап сбора данных из магазинов. Время идет, технологии сбора информации также развиваются. Если раньше для получения данных об объеме продаж специалисты компании собирали эти данные вручную, посещая магазины и заполняя соответствующие бланки специальной формы, записывая название артикула, количество товара на складе, количество

закупок по документам доставки и прочее, то в настоящее время процесс сбора данных в значительной мере автоматизирован. При «ручном» сборе процесс слишком затягивается, есть вероятность ошибки при записи артикула либо неверной записи количества на складе, закупленного количества. Особенно при большом ассортименте возрастает влияние человеческого фактора на качество данных. В разных компаниях («ACNielsen»¹, «GfK Group»²) используются различные автоматизированные системы данных.

В «GfK Retail and Technology» в России тоже не всегда процесс сбора данных был автоматизирован. В 1995 году, когда исследование стартовало в России, не все розничные компании имели автоматизированные системы учета продаж и складских остатков. Приходилось работать с бухгалтерской отчетностью, а также с внутренними записями компании, так как бухгалтерская отчетность не всегда содержала необходимую детализацию, как указано в структуре данных. Даже несколько лет назад еще существовали компании, где сбор данных был частично автоматизирован: после выгрузки данных из внутренней системы учета служащим компании приходилось их дополнять вручную. Сейчас же в компанию поступают только электронные данные строгой структуры как из головных офисов сетевых компаний-ритейлеров, так и независимых небольших компаний, которые, тем не менее, используют современные системы автоматизированного учета, такие, как например 1С. После поступления данных в «GfK» они загружаются в специальную систему компании. Данные проходят проверку на соответствие установленной структуре, набору торговых точек, товарных групп и других важных параметров (табл. 2). Главная задача на этом этапе — обработать объем входящих данных, очистить от ошибок и привести в вид, удобный для последующей обработки.

Таблица 2

Пример структуры данных, поступающих от ритейлера

Магазин (код)	Категория	Товарная группа		Подгруппа	Марка	Артикул	Цена, руб.	Кол-во, шт.
		наимен.	код					

Источник: GfK workflow

После этого данные кодируются в соответствии с международным кодовым планом, принятым в компании по различным характеристикам. Далее они проходят еще многоступенчатую проверку и обработку — логический и содержательный контроль — и на их основании формируются отчеты для пользователей (клиентов-производителей техники и розничных компаний — участников исследования).

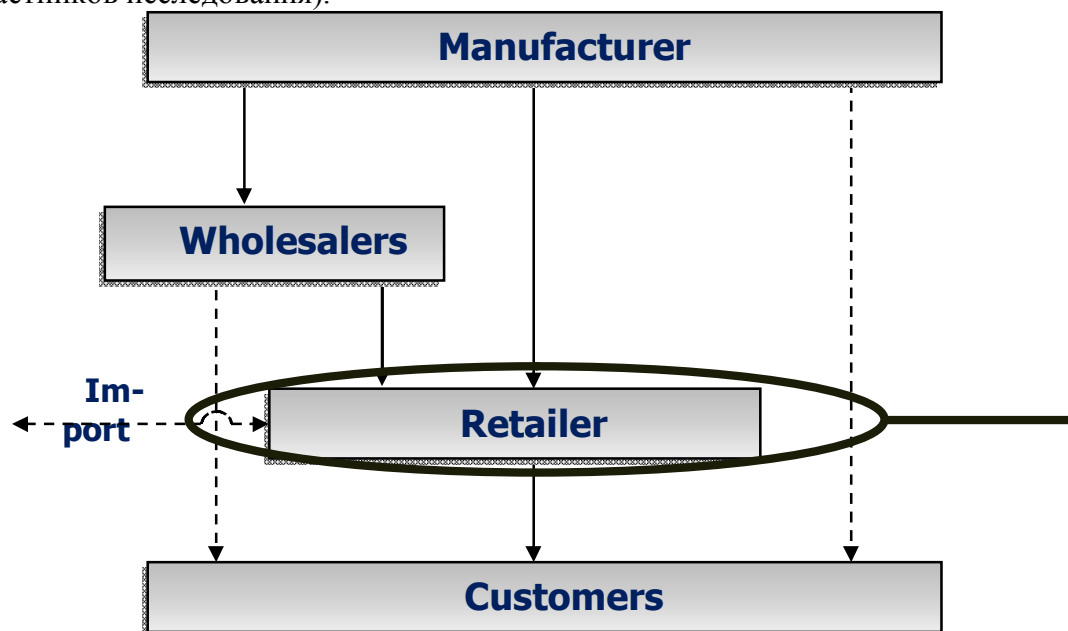


Рис. 2. Схема получения данных от компаний «GfK Group».

¹ URL: <http://www.acnielsen.ru/> (дата обращения - 25.01.11)

² URL: <http://www.gfk.com/> (дата обращения - 25.01.11)

Эффективность этого метода обуславливается следующими факторами:

- 1) получение данных о продажах напрямую из базы учета продаж ритейлера (отсутствие искажения);
- 2) заинтересованность ритейлера в предоставлении данных (взамен получает ритейлерский отчет, который содержит общее описание рынка, его характеристики и динамику за последний период с указанием доли агента на рынке);
- 3) научно-обоснованная методология, позволяющая получать репрезентативные данные из нужных каналов для создания представления о рынке в целом;
- 4) налаженная система подбора магазинов, сбора и обработки данных.

В настоящее время клиентам «GfK Retail and Technology» доступна ежемесячная и еженедельная подписка на отчеты по более чем 300 недовольственным товарным категориям, среди которых:

- аудио-видео техника;
- компьютерная и офисная техника;
- телекоммуникация;
- крупная и малая бытовая техника;
- фототовары;
- товары для сада, ремонта, строительства;
- оптика медицинского назначения;
- автомобильные шины, масла, запчасти и многие другие.

Как видно из этого далеко не полного списка укрупненных категорий, метод компании «GfK» применим и успешен практически для любого розничного рынка товаров и услуг.

Применение нейронауки в социологическом исследовании

Сланевская Нина Михайловна

Центр междисциплинарной нейронауки, Санкт-Петербург

Предметом социальной когнитивной нейронауки является взаимодействие факторов, принадлежащих к трем уровням — социальному, когнитивному и нейронному. Социальный уровень соответствует мотивационным и социальным факторам, влияющим на поведение и опыт, когнитивный уровень — механизмам обработки информации, способствующим возникновению явлений на социальном уровне, нейронный уровень соответствует механизмам мозга, отвечающим за когнитивный процесс¹. Когнитивный уровень — это уровень, где нейронаука и социальные науки взаимодействуют больше всего. Достижения современной социальной когнитивной нейронауки могут значительно обогатить социологическую методологию.

Принято считать, что голосование за политического лидера — это рациональный выбор голосующего, который надеется, что кандидат принесет материальные выгоды и выразит ценности голосующего (когнитивно контролируемый процесс). При таком процессе обычно активируется, главным образом, префронтальная латеральная кора полушарий мозга, отвечающая за выполнение решений. Однако возможно, что голосующий предпочтет просто красиво выглядящего кандидата (аффективно автоматический процесс), что предполагает работу лимбической части мозга, отвечающую за эмоции.

Если при двух разных социологических опросах на одну и ту же тему (с одним и тем же опрашиваемым), вопросы сформулированы так, что человек при ответе должен основываться на разный тип мышления: сознательное рассуждение в одном и бессознательное привычное в другом, то исследователь может получить противоречивые результаты.

При исследовании мозга было выявлено, что амигдала (структура мозга) играет важную роль, в основном, в негативной реакции. Сканирование мозга помогает выявить внутреннюю негативную реакцию, даже если участник эксперимента отрицает свое негативное предубеждение в открытом сознательном опросе. Так, например, было выявлено поведенческое и нейро-

¹ Ochsner K., Lieberman M. The Emergence of Social Cognitive Neuroscience // American Psychologist. 2001. Vol. 56. No. 9. P. 717–734.

биологическое несоответствие при изучении восприятия лиц разных расовых групп. Сканирование мозга при одновременном рассмотрении фотографий людей другой расы показало активацию амигдалы, что означает отрицательные эмоции и настороженность, а в ответах на вопросы об отношении к другой расе участники говорили о своем положительном или нейтральном отношении¹.

Активация базальной ганглии показывает, наоборот, нашу благосклонную предрасположенность к чему-то. Ее активация происходит, например, когда смотришь на фото любимого человека (или любимого политика). А активация базальной ганглии вместе с допаминергической нейротрансмиттерной системой (химическое вещество мозга, связанное с положительными эмоциями) может являться мотивационным источником сама по себе и поэтому может заставить сделать бессознательный выбор².

Либерман (Lieberman) предлагает теорию о рефлексивной (reflective) С-системе и рефлексивной (reflexive) Х-системе³. Х-система, состоящая из латеральной височной коры, амигдалы и базальной ганглии (с двух сторон каждая) часто спонтанно и бессознательно интегрирует текущие задачи, контекст и восприятие в связанное целое. С-система, состоящая из префронтальной коры, передней части поясной извилины и медиальной височной доли (с двух сторон каждая), призывается к работе, если Х-системе не удастся создать связную информацию на выходе. Эти системы различаются по модели работы. Х-система постоянно интегрирует информацию из многих источников одновременно с большой скоростью и эффективностью и зависит, в основном, от ассоциативных связей, которые образовались в результате длительного обучения или опыта. С-система может обрабатывать только несколько единиц информации за один раз, но зато может произвольно конструировать ассоциации между единицами информации, как требуется в данном контексте. Ее эффективность зависит от мотивационных факторов человека, решившего использовать сознательные ресурсы, которыми он владеет в данный момент. Процессы в С-системе в большей степени доступны сознанию: информация обрабатывается последовательно, а не параллельно как в Х-системе, и организуется языковыми средствами. В Х-системе информация обрабатывается одновременно и параллельно, процесс менее доступен сознанию, и в нем не используются символы, как в С-системе. Если политический опросник отдает предпочтение только одной из двух систем (Х- или С-системе), то экспертные знания будут искажены и не полны. Если вопрос ассоциируется со страхом у отвечающего, то страх приведет к преобладающему использованию Х-системы. Вопросы, относящиеся к оценке справедливости, могут заставить снизить порог перехода из Х-системы в С-систему, и С-система начинает работать и старается найти «справедливое решение». Установившиеся нейронные связи при заучивании повторяемого социального опыта трудно переделать, поэтому надо учитывать то, как свойства нейронных связей влияют на поведение и когнитивные способности при составлении опросников. Нейронные связи, выявленные при сканировании, могут дать значимую информацию для предсказания поведения в реальном мире.

Успех принятия решений в социальных условиях зависит от нашей способности понимать намерения, эмоции и мнения других. У человека есть особые нейронные механизмы, которые обеспечивают эту функцию в той или иной мере. Это система зеркальных нейронов, которая помогает понять моторные движения и намериваемые действия других. Есть также способность, под названием эмпатия, которая помогает понять и разделить эмоции и сенсорные ощущения других. И третья составляющая нашего нейронного механизма общения — это «теория о мышлении», которая позволяет нам понимать абстрактные концепции, верования и желания других людей; «теория о мышлении» говорит о том, как думают другие люди.

Исследования в нейронауке показывают, что у человека есть определенные нейронные механизмы, которые задействованы для понимания как своих, так и чужих ментальных состояний.

¹ Phelps E., Thomas L. Race, Behaviour, and the Brain: The Role of Neuroimaging in Understanding Complex Social Behaviours // *Political Psychology*. 2003. Vol. 24. No. 4. P. 747–758.

² Dupue R.A., Collins P.F. Neurobiology of the Structure of Personality: Dopamine, Facilitation of Incentive Motivation, and Extraversion // *Behavioural and Brain Sciences*. 1999. Vol. 22. P. 491–569.

³ Lieberman M.D., Schreiber D., Ochsner K. Is Political Cognition Like Riding a Bicycle? How Cognitive Neuroscience Can Inform Research on Political Thinking // *Political Psychology*. 2003. Vol. 24. No. 4. P. 681–704.

Эти общие характерные нейронные сети для всех модулируются в определенном социальном контексте, при оценке этого контекста и в общении с другим человеком. Крис Фрит и Таня Зингер рассматривают четыре области, где есть такие взаимодействия: наше ощущение справедливости, альтруистическое наказание, доверие и рамочный эффект¹. Рамочный эффект — это взаимодействие эмоций с рациональным решением, при котором, одна и та же проблема, сформулированная по-разному (рамочная формулировка) ведет к разному решению. Рамочный эффект базируется на эмоциях, которые появляются в зависимости от формулировки проблемы.

Потребность в социальных связях является сильной мотивацией для человека и руководит его действиями при контакте с лицами, которых он уважает, власть которых над собой признает, в семейных отношениях, на работе и с теми, кого он любит². Обычно, когда люди видят любимого или любимую, то активируются базальные ганглии. Социальная поддержка влияет на сердечнососудистую систему: систологическое и диастолическое давление крови повышается гораздо меньше при стрессовой ситуации, если есть социальная поддержка³. Недостаток социальной поддержки гораздо более серьезный риск, ведущий к увеличению заболеваемости и смертности, чем курение, переедание или высокое кровяное давление⁴.

Важная роль принадлежит ощущению справедливости в социальной среде. Справедливость и сотрудничество — две прямо взаимозависимые категории⁵. Несмотря на то, что деньги и другие материальные ценности без сомнения приносят удовлетворение, последние нейронаучные исследования показывают, что такие нематериальные факторы как равенство, справедливое решение и справедливое наказание, могут принести даже большее удовлетворение. Нейронаучные исследования показали, что справедливые предложения («игра в доверие») вели к ощущению большего счастья и увеличивали активность нескольких областей мозга, известных как «нейросети удовольствия и награды», в отличие от несправедливых предложений⁶. Активировались «нейросети удовольствия и награды» тогда, когда участники игры наказывали нечестных игроков, жертвуя своим собственным денежным вознаграждением только для того, чтобы восторжествовала справедливость. Такие результаты нейронаучного исследования очень важны для социальной организации людей.

Представляется, что привлечение технологий нейронауки — сканирование мозга в реальном времени с экспериментальной задачей социолога — или использование готовой нейронаучной базы данных социальных нейронаук (нейроэкономика, нейросоциология, нейрополитология, нейротеология, нейроморальность и нейроэтика) и нейронаучных теоретических объяснений обогатило бы методологическую базу социологии.

Новый способ интерпретации результатов анализа соответствий

Шафир Марк Александрович

Исследовательская компания «РАДАР», НИУ ВШЭ

1. Введение

Сегодня трудно недооценить роль анализа соответствий⁷ (АС) в социологических и, особенно, в маркетинговых исследованиях. Этот метод часто используется для построения карт вос-

¹ Frith Ch., Singer T. The Role of Social Cognition in Decision-making // Philosophical Transactions of the Royal Society B. 2008. Vol. 363. P. 3875–3886.

² Lieberman M. Social Cognitive Neuroscience: A Review of Core Processes // The Annual Review of Psychology. 2007. Vol. 58. P. 259–289.

³ Lepore S.J., Allen K.A., Evans G.W. Social Support Lowers Cardiovascular Reactivity to an Acute Stressor // Psychosomatic Medicine. 1993. Vol. 55. P. 518–524.

⁴ Eisenberger N., Taylor S., Gable S., Hilmert C., Lieberman M. Neural Pathways Link Social Support to Attenuated Neuroendocrine Stress Responses // NeuroImage. 2007. Vol. 35. P. 1601–1612.

⁵ Hegtvedt K.A., Killian C. Fairness and Emotions: Reaction to the Process and Outcomes of Negotiations // Social Forces. 1999. Vol. 78. P. 269–303.

⁶ Tabibnia G., Lieberman M. Fairness and Cooperation are Rewarding: Evidence from Social Cognitive Neuroscience // Annals of the New York Academy of Sciences. 2007. Vol. 1118. P. 90–101.

⁷ Анализ соответствий (Correspondence analysis) — разведочный метод анализа данных, позволяющий визуализировать любую таблицу сопряженности в виде карты в пространстве низкой размерности, где строки и столбцы

приятия при анализе имиджа брендов, при описании сегментов, полученных с помощью кластерного анализа, и вообще, для анализа любых таблиц сопряженности. Анализ соответствий снискал огромную популярность благодаря своей способности эффектно визуализировать большие массивы данных.

Однако за кажущейся простотой интерпретации результатов метода таится скрытая угроза: при поверхностном взгляде на карту соответствий можно получить неправильное представление о взаимосвязи точек и таким образом прийти к неверным выводам. Суть этой угрозы кроется в том, что анализ соответствий изначально разрабатывался как разведочный, «эвристический»¹ метод анализа данных; в нем не заложен аппарат тестирования гипотез² и, следовательно, отсутствует механизм переноса результатов с выборки на генеральную совокупность. Это, в свою очередь, затрудняет интерпретацию результатов метода, делая ее многозначной, так что один и тот же результат может быть истолкован по-разному. Если учесть, что анализом соответствий подчас пользуются люди, не знакомые с ним, соблазнившиеся его обманчивой простотой³, то становится понятным, что этот метод в «неумелых руках» может из эффективного аналитического метода превратиться в инструмент принятия неправильных решений.

В настоящем выступлении представляется авторский подход к интерпретации результатов анализа соответствий. Данный подход позволит аналитику проинтерпретировать числовые коэффициенты АС (вклады и квадраты корреляций) по схеме, максимально приближенной к схеме интерпретации результатов метода главных компонент (разведочного факторного анализа), а также построить карту, понятную любому человеку, вероятность неправильного толкования которой будет сведена к минимуму. Основные принципы предлагаемого подхода следующие:

- 1) упор на сходство анализа соответствий с методом главных компонент, достигающееся использованием восстановленных коэффициентов корреляции точек с осями;
- 2) разграничение карты соответствий на сектора в соответствии со значениями квадратов косинусов.

Данный подход был разработан автором в ходе семилетней практической деятельности и апробирован в преподавании⁴, в ряде научных публикаций⁵ и выступлениях на научных семинарах и конференциях.

2. О формате основных результатов анализа соответствий

Для дальнейшего изложения необходимо дать короткое пояснение формату результатов анализа соответствий, которые обычно выдаются статистическими пакетами. К слову, анализ соответствий реализован в популярнейших пакетах SPSS (лучше через синтаксис), Statistica (через меню), SAS (только через синтаксис). В стандартный «аутпут» анализа соответствий входят следующие результаты:

1. Таблица *сингулярных чисел по осям* и соответствующих этим числам процентов объясненной информации; используется, чтобы определить, какой % исходной информации содержится в двумерном решении, и имеет ли смысл вводить в анализ дополнительные оси.
2. *Координаты* точек-строк и точек-столбцов; на основе координат строится карта соответствий.
3. *Двумерная карта соответствий* в выбранных осях (обычно по умолчанию выводятся оси №1 и 2, но можно смотреть карты и в других осях, например, 1 и 3 и т.д.). Служит основной для интерпретации результатов АС. Может быть симметричной (точки-строки и точки-столбцы

таблицы будут представлены в виде точек. Разработал метод Ж.-П. Бензекри, см. классическую работу: Benzécri J.-P. Analyse des données. Tome 2 : L'analyse des correspondences. Paris: Dunod, 1973.

¹ По выражению А.О. Крыштановского.

² Отсутствие аппарата тестирования гипотез рассматривалось создателем метода как преимущество; Бензекри намеренно создавал метод «свободный от модели», чтобы «модель вытекала из данных, а не наоборот».

³ Это часто происходит в сфере маркетинговых исследований в связи с нехваткой профессиональных кадров.

⁴ Автор преподает данный подход к АС магистрантам НИУ ВШЭ в рамках курса «Современные методы анализа данных», см. подробнее страницу автора на портале НИУ ВШЭ: <http://www.hse.ru/org/persons/506080>

⁵ Шафир М.А. Методологические истоки анализа соответствий // Социология: 4М. 2008. №28. С. 29–44; Шафир М.А. Использование анализа соответствий для изучения имиджа продукта: на примере анализа имиджа московских банков // Методы социологических исследований. М.: ТЕИС, 2006.

относительно равномерно распределяются на плоскости) и ассиметричной¹ (точки-столбцы образуют вершины, а точки-строки находятся внутри них; или наоборот).

4. Набор *вкладов* точек-строк и точек-столбцов в каждую ось (или только в выбранные оси). Вклады показывают, какой процент инерции оси объясняется данной точкой, иначе говоря, в какой мере данная точка влияет на геометрическую конфигурацию оси. Традиционно вклады являются основными показателями при интерпретации осей в АС. В методе главных компонент аналогом этого показателя являются так называемые факторные значения (*factor scores*²), крайне редко используемые при интерпретации результатов.

5. Набор *квадратов корреляций* (*квадратов косинусов*) точек-строк и точек-столбцов с каждой из осей (или только с выбранными осями). Квадраты корреляции, как понятно из названия, показывают, насколько точка содержательно связана с осью. Альтернативное название показателя, квадраты косинусов, отсылает к геометрическому представлению данных и опять же, как следует из названия, показывает угол наклона между вектором точки и осью. В анализе соответствий традиционно используется как мера качества представленности точки в контексте выбранной оси. В методе главных компонент аналогом квадратного корня из этого показателя служат факторные нагрузки (*factor scores*³).

6. *Суммы квадратов корреляций* каждой точки по выбранным осям. Традиционно является мерой качества представленности точки во всем решении. В методе главных компонент аналогом этого показателя являются «общности» (*communalities*)⁴.

3. Недостатки существующих подходов к интерпретации анализа соответствий

С момента появления анализа соответствий различные авторы, начиная от самого основателя метода Ж.-П. Бензекри и его ученика М. Гринакра⁵ до современных российских исследователей А. Кутлалиева и А. Попова⁶, предлагали свои подходы к интерпретации результатов анализа соответствий, в т.ч. многочисленные приемы и хитрости (например, использование дополнительных точек, введение 3-й оси или осей еще более высокого порядка и т.д.). Большинство из этих подходов повторяет друг друга и оригинальные работы Гринакра и сосредотачивается вокруг поиска т.н. «объясняющих» и «объясняемых»⁷ точек, которые имеют значения вкладов и квадратов корреляций выше среднего (или выше иного порога). Эти подходы мало что дают для визуализации результатов на карте (правда, количество точек, показываемых на карте, уменьшается) и не снижают рисков неправильной интерпретации карты. Главная опасность симметричных карт состоит в том, что расстояния между точками-строками и точками-столбцами напрямую сравнивать нельзя, так как они не определены в методе. Вместо этого, как говорится во всех работах по АС, нужно интерпретировать относительную позицию одной точки-строки по отношению ко всем точкам-столбцам, или наоборот. Иначе говоря, нужно сначала проинтерпретировать оси (назвать полюса осей), а потом сравнивать положение каждой конкретной точки с этими осями.

Это и есть главный камень преткновения при интерпретации карты соответствий для человека, не знакомого с этим ограничением. Сравнивая эти расстояния напрямую, можно получить неправильное представление о структуре взаимосвязей между переменными. Иллюстрацию подобной ошибки можно увидеть на рис. 1.

¹ Подробнее об ассиметричных картах см.: Greenacre M.J. Correspondence Analysis and its Interpretation // Correspondence Analysis in the Social Sciences. San Diego, CA: Academic Press. 1994. P. 3–22.

² Иначе говоря, это — коэффициенты уравнений, показывающие, на какую величину нужно умножить каждую из переменных, на которых строится фактор, чтобы получить значение фактора.

³ Коэффициенты корреляции переменной и фактора, использующиеся для содержательной интерпретации факторов.

⁴ Процент дисперсии исходной переменной, сохранившийся при переходе в пространство сниженной размерности.

⁵ Greenacre M.J. Correspondence analysis in practice. London: Academic Press, 1993.

⁶ Кутлалиев А., Попов А. Эффективность рекламы. М.: ЭКСМО, 2005.

⁷ По Кутлалиеву и Попову (цит. соч.), *объясняющие* (экспликативные) точки — это точки, формирующие определенную ось; их вклад в главную инерцию выше, чем вклады других точек. *Объясняемые* точки — точки, которые могут не вносить вклад в формирование оси, но прилегают к ней (угол между направлением на точку и осью достаточно мал).

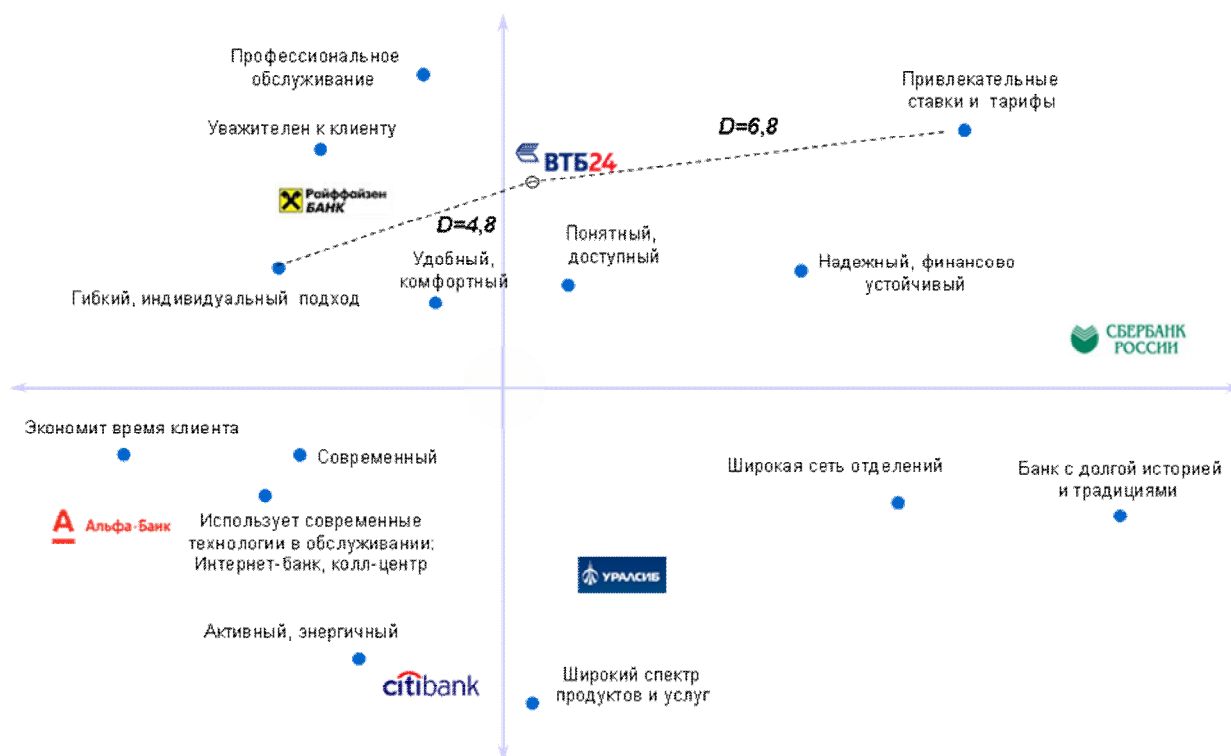


Рис. 1. Ошибка при сравнении расстояний между точками из разных наборов на двумерной карте соответствий.

На приведенной карте соответствий расстояние от банка ВТБ24 до атрибута «привлекательные ставки и тарифы» составляет 6,8¹. Расстояние от того же банка до атрибута «гибкий, индивидуальный» значительно меньше и составляет 4,8. Логично было бы сделать вывод, что ВТБ24 в большей степени «индивидуальный», чем «с привлекательными тарифами», что ошибочно. На самом деле, ВТБ24 обладает атрибутом «привлекательные ставки и тарифы», но не обладает атрибутом «гибкий, индивидуальный подход»², т.к. ВТБ24 имеет практически нулевой квадрат корреляции с горизонтальной осью, и расстояние по горизонтали для него не имеет никакого значения; главное — расстояние по вертикали.

Без значений квадратов корреляции «под рукой» мы бы не увидели этого и могли бы сделать ошибочный вывод. Заказчики исследований редко имеют перед глазами таблицу квадратов корреляций, а даже если она есть в приложении к отчету, вряд ли будут в нее смотреть. Им хочется видеть все на карте соответствий, что, в общем, логично: если мы говорим, что метод славится эффектной визуализацией, то хотелось бы наблюдать эту визуализацию «на одном слайде». Поэтому представляется закономерным, что аналитик должен заранее проанализировать все данные и представить такую карту соответствий, которая бы легко считывалась и приводила к понятным всем выводам. Очевидно, что для этой цели нужно перерабатывать карты соответствий, выдаваемые программами.

Чтобы избежать ошибок с неправильной интерпретацией расстояний между точками, М. Бендиксен предлагает следующий подход: убрать с карты все точки-строки, и оставить только точки-столбцы, или наоборот, заменить убранный набор названием полюсов осей. Далее интерпретируется попадание бренда в тот или иной квадрант. Проинтерпретировав полюса осей, и убрав с карты все атрибуты, мы можем получить следующую картину:

¹ Расстояние в данном случае является безразмерной величиной.

² Это легко подтверждается анализом стандартизованных остатков. Так, значение стандартизованного на пересечении «ВТБ24» и атрибута «привлекательные ставки и тарифы» составляет 2,14; а аналогичное значение для атрибута «гибкий, индивидуальный подход» составляет 0,71. При пороговом значении 1,96 (соответствует доверительной вероятности 95%) гипотезу об отсутствии связи между банком и первым атрибутом следует отклонить, а для второго атрибута принять.

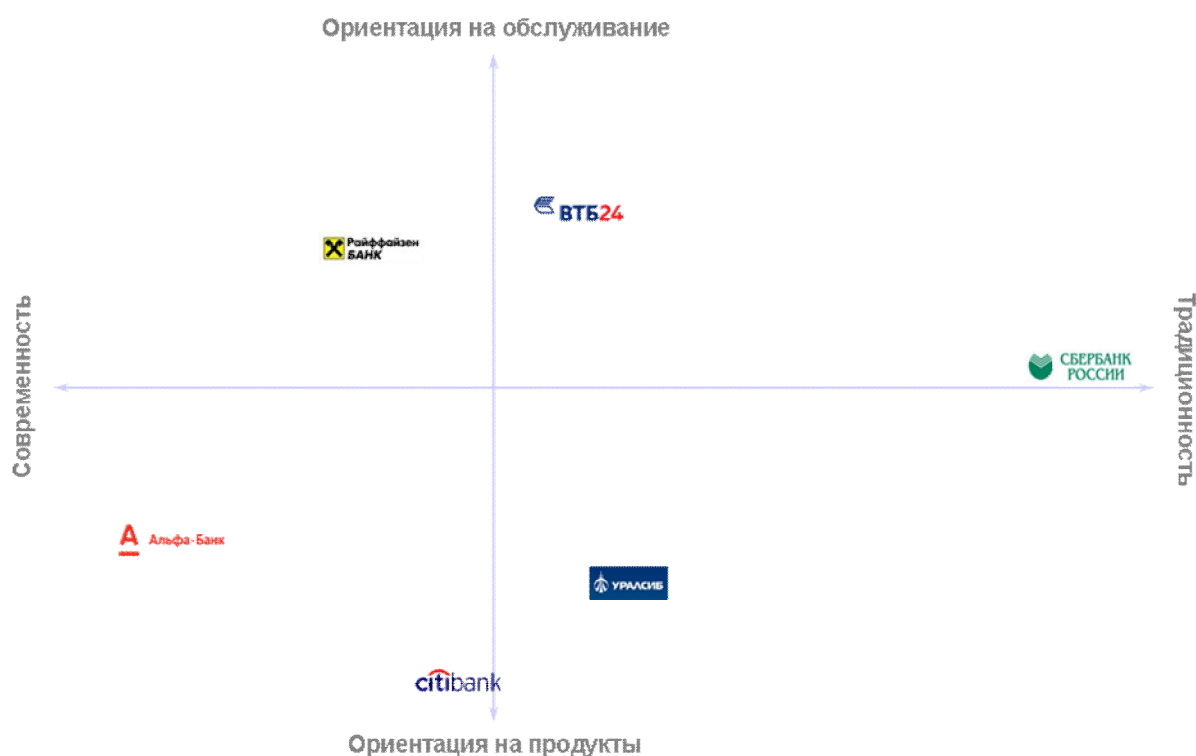


Рис. 2. Альтернативный способ интерпретации карт соответствий:
по объясняющим точкам называются полюса осей,

затем объясняющие точки удаляются с карты, и остаются только объясняемые точки.

Однако, взглянув на подобную карту, заказчики исследований снова недоумевают: почему, если преимуществом анализа соответствий является эффектная визуализация, нужно убирать половину данных? И это, в принципе, снова резонный вопрос.

Наконец, еще один недостаток традиционного подхода к интерпретации результатов АС, предлагаемый в большинстве работ, — это интерпретация осей, основанная исключительно на абсолютных вкладах. Согласно данному подходу, чтобы назвать оси / полюса осей, мы должны пользоваться показателями абсолютных вкладов, которые есть не что иное как процент инерции оси, объясняемый конкретной точкой. Это довольно грубый показатель, который, к тому же, не очень удобен на практике, т.к. значения вкладов точек по одной оси сходятся к 100%, и при большом числе точек могут быть очень маленькими, что затрудняет их сравнение.

Если вспомнить метод главных компонент, то можно привести следующую аналогию: при назывании фактора пользоваться не привычными факторными нагрузками (factor loadings), показывающими корреляцию между переменной и фактором, а факторными коэффициентами (factor scores), показывающими, на какое значение нужно умножить переменную, чтобы получить искомый фактор. Последняя таблица даже не выводится по умолчанию в SPSS, ее нужно выводить вручную, что говорит о том, что создатели SPSS посчитали (и не без оснований) данную таблицу не очень важной при интерпретации результатов метода главных компонент.

Есть куда более «полезный» показатель — квадрат корреляции между точкой и осью. На наш взгляд, данному показателю следует уделить более пристальное внимание.

4. Сходство анализа соответствий с методом главных компонент

Многие авторы подчеркивают сходство анализа соответствий с методом главных компонент¹. Действительно, и в том, и в другом методе к исходной матрице применяются линейные преобразования; оба метода, по сути, являются методами снижения размерности; и в том, и в другом методе объясняемая информация перераспределяется между осями и упорядочивается так, что первые оси объясняют наибольший процент исходной информации и т.д. Математиче-

¹ См., например: Greenacre M.J., Blasius J. Computation of Correspondence Analysis // Correspondence Analysis in the Social Sciences. San Diego, CA: Academic Press. 1994. P. 53–78.

ская основа анализа соответствий — сингулярное разложение — является, по сути, общим случаем разложения по собственным значениям, используемого в методе главных компонент¹. Более того, часть показателей в АС, как было показано выше, имеет аналоги в методе главных компонент. Тогда возникает вопрос: почему бы в анализе соответствий не использовать общую логику интерпретации метода главных компонент? Если это сделать, то значительно облегчается процесс интерпретации осей (называния полюсов осей), что, в итоге, позволит получить ряд выводов, даже не глядя на карту.

Однако в «природе» (т.е. в математической подоплеке) двух методов заложено существенное различие. Если в методе главных компонент название фактора отражает наличие или отсутствие какого-либо явления (например, «склонность к риску», «обеспеченность», «новаторство», и т.д.), то в анализе соответствий обычно называют оба полюса оси, т.к. они могут быть содержательно разными².

В любом случае для приближения интерпретации анализа соответствий к интерпретации метода главных компонент нужно получить матрицу, аналогичную матрице факторных нагрузок. Последняя представляет собой таблицу корреляций между всеми переменными и выделенными в решении факторами. В данной таблице обычно сортируют значения по убыванию и убирают коэффициенты корреляции, меньше по модулю чем 0,3 или 0,4³.

Коэффициенты корреляции между точками и осями в АС можно получить, если из квадрата корреляции извлечь квадратный корень и приписать знак координаты точки (т.к. сам по себе квадрат корреляции не имеет знака). Возьмем таблицу квадратов корреляций для точек-строк (имиджевые атрибуты) по первым двум осям:

	1	2
Активный, энергичный	0,21	0,72
Надежный, финансово устойчивый	0,81	0,13
Понятный, доступный	0,17	0,43
Профессиональное обслуживание	0,32	0,59
Удобный, комфортный	0,16	0,42
Гибкий, индивидуальный подход	0,68	0,20
Широкая сеть отделений	0,79	0,07
Уважителен к клиенту	0,44	0,43
Привлекательные ставки и тарифы	0,38	0,48
Использует современные технологии: Интернет-банк, колл-центр	0,84	0,11
Банк с долгой историей и традициями	0,90	0,04
Широкий спектр продуктов и услуг	0,05	0,81
Современный	0,67	0,04
Доставляет радость	0,56	0,10
Экономит время клиента	0,74	0,02

С помощью функций «корень» и «знак», а также после сортировки и применения условного форматирования в MS Excel, легко получить следующую таблицу:

	1	2
Банк с долгой историей и традициями	0,95	
Надежный, финансово устойчивый	0,90	
Широкая сеть отделений	0,89	
Привлекательные ставки и тарифы	0,62	0,69
Понятный, доступный	0,42	0,65

¹ Eckart C., Young G. The approximation of one matrix by another of lower rank // Psychometrika. 1936. Vol. 1. P. 211–218.

² Это различие проистекает из вида данных, подающихся на вход метода. Метод главных компонент создан для работы с интервальными переменными, значения которых строго упорядочены, а анализ соответствий — для работы с номинальными, значения которых являются категориями и не могут быть упорядочены сами по себе.

³ Такой подход рекомендуется: Бююль А., Цефель П. SPSS: искусство обработки информации. СПб.: ДиаСофт, 2005.

Широкий спектр продуктов и услуг		-0,90
Удобный, комфортный	-0,41	0,65
Активный, энергичный	-0,46	-0,85
Профессиональное обслуживание	-0,56	0,77
Уважителен к клиенту	-0,66	0,66
Гибкий, индивидуальный подход	-0,82	0,45
Современный	-0,82	
Экономит время клиента	-0,86	
Использует современные технологии: Интернет-банк, колл-центр	-0,92	

Теперь мы имеем таблицу корреляций точек с осями, которую намного проще интерпретировать. Очевидно, что положительное направление первой оси образуют атрибуты, связанные с традиционностью и следующими из нее надежностью и масштабностью. Отрицательное направление первой оси — это, напротив, современность и использование современных технологий. Положительное направление второй оси образуют точки, объединенные темой ориентированности на клиента, простоты, доступности (в т.ч. финансовой); тогда как отрицательный полюс говорит об ориентации скорее на конкретные продукты и услуги, на широту их спектра; на активность банка в освоении новых услуг и технологий.

5. Разграничение карты соответствий на сектора по значениям квадратов косинусов

Проинтерпретировав таким образом оси, можно перейти к работе с картой. Чтобы облегчить восприятие карты людям, не знакомым с понятиями вкладов и квадратов корреляций, и сделать карту самодостаточной, мы предлагаем описанный ниже прием.

По сути, квадрат корреляции точки и оси является квадратом косинуса угла, образованного вектором точки и осью. Чем больше косинус, тем сильнее точка «притягивается» к оси, лежит в ее направлении. Если косинус угла равен 1, то точка вообще сливается с осью, лежит на ней. Для иллюстрации приведем следующий рисунок, на котором показаны значения косинуса и синуса на окружности:

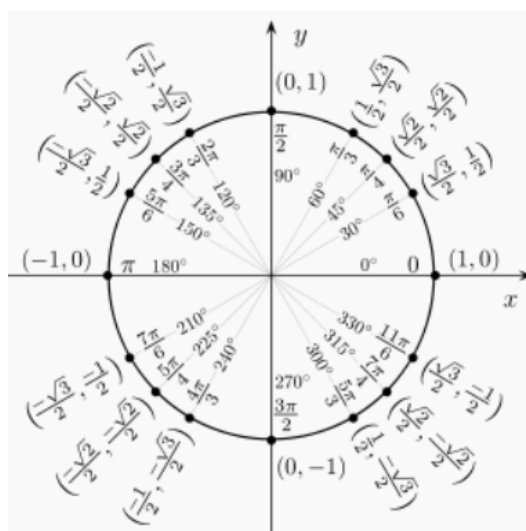


Рис. 3. Значения косинуса и синуса углов на окружности.

Примечательным является угол в 45° , т.е. биссектриса. Любой угол меньше 45° образует вектор, лежащий ниже биссектрисы, т.е. более «притянутый» к оси X. Угол больше 45° образует вектор, лежащий выше биссектрисы, т.е. более «притянутый» уже скорее к оси Y. Косинус угла в 45° равен $\sqrt{2}/2$, а его квадрат – ровно $1/2$. Квадрат косинуса, равный $0,5$, означает, что половина инерции данной точки принадлежит к данной оси, а оставшаяся половина инерции распределена между другими осями, и не больше, чем на данной оси¹.

¹ Кутлалиев А., Попов А. Эффективность рекламы. Москва: ЭКСМО, 2005.

Отсюда вытекает простой вывод: нужно изобразить на плоскостной карте соответствий все четыре биссектрисы. На биссектрисах углов могли бы лежать точки, чей квадрат корреляции равен 0,5 для двух первых осей (в реальности такие точки встречаются очень редко). Соответственно, точки, лежащие между двумя соседними биссектрисами будут иметь схожие значения квадрата корреляции с обеими осями.

Таким образом, если исключить точки, лежащие в самом центре карты (они, как правило, имеют нулевые квадраты корреляции с обеими осями, и, таким образом, не объясняются выбранными осями вообще, так что их следует исключить из интерпретации), то можно говорить о наличии четырех секторов на карте. Мы не будем напрямую сравнивать расстояния между точками-строками и точками-столбцами, но получаем возможность говорить о взаимосвязи точек, попавших в один сектор; здесь взаимосвязь понимается как схожесть квадратов корреляций точек с осями, что, в принципе, не противоречит классической интерпретации АС. Пример карты, построенной таким образом, представлен на рис. 4.

Если ограничить интерпретацию положения точек рамками нарисованных секторов, то она становится понятнее. Чтобы определить, какие атрибуты связаны с какими банками, можно руководствоваться следующей логикой: связанными можно назвать точки, попавшие в один сектор. При этом точки, лежащие на биссектрисах или близко к ним, могут в равной степени относиться к соседним секторам.



Рис. 4. Интерпретация карты соответствий с выделением секторов.

Рассмотрим для примера верхний сектор «клиентоориентированности». Его образуют такие высказывания, как «профессиональное обслуживание», «уважителен к клиенту», «удобный, комфортный», «понятный, доступный», «привлекательные ставки и тарифы». В этот сектор попали ВТБ24 и Райффайзенбанк, которые, как следует из данной карты, имеют схожий имидж клиентоориентированного банка. Однако при этом Райффайзенбанк находится практически на биссектрисе, что говорит также о его современности. То же самое можно сказать о высказывании «привлекательные ставки и тарифы»: оно в равной степени относится и к верхнему сектору, и к правому, где находится Сбербанк, но мало подходит Райффайзенбанку, т.к. угол между этими двумя векторами уже около 90° ; они связаны, но очень слабо.

Можно изображать не четыре сектора, а больше, деля каждый из них на новые сектора, не обязательно равные. Чем уже будет сектор, тем сильнее будут связаны точки, в него попавшие. В этом легко убедиться с помощью анализа стандартизованных остатков.

6. Итоги

Автором представлен новый подход к интерпретации результатов анализа соответствий, основанный на приближении к интерпретации метода главных компонент в части чисел, и на анализе углов и образуемых ими секторов в части графического решения.

Мы считаем данный подход более корректным и более эффективным как со стороны аналитика, так и со стороны заказчика. Аналитик интерпретирует результаты, используя понятный, наглядный и содержательно осмысленный показатель корреляции; заказчик же получает самодостаточную карту, на которой уже выделены секторы взаимосвязанных переменных.

Для проверки результатов анализа соответствий рекомендуется также использовать анализ стандартизованных остатков.

Он-лайн методы в качественных исследованиях

Шейхетов Сергей Владимирович, *Ipsos*

Словосочетание «маркетинговые исследования он-лайн» стало чрезвычайно популярным в последнее время. Многочисленные публикации в различных около-маркетинговых изданиях растиражировали этот термин до такой степени, что сформировался ряд стереотипов, связанных с исследованиями в Интернете.

Во-первых, он-лайн исследования преподносятся как, безусловно, прогрессивные и современные, уже считается неприличным для уважающего себя агентства не проводить, или, по крайней мере, не заявлять о наличии подходов для исследования он-лайн.

Во-вторых, он-лайн методы, почему-то противопоставляются традиционным исследовательским методам. Причем, в любой статье, в любой презентации на конференции приличным считается сказать, что да, вот де, мол, существуют новые, перспективные и интересные методы он-лайн, но они никогда не станут альтернативой face-to-face интервью или старым добрым фокус-группам. Причем, критики редко принимают во внимание тот факт, что те специалисты, которые действительно проводят исследования в сети, никогда не противопоставляют он-лайн и офф-лайн, более того, почти всегда настаивают на комбинировании двух подходов.

В своем выступлении я хотел бы показать, что в этих стереотипах соответствует действительности, а что нет. Для этого я отвечу на вопросы: почему именно он-лайн? Для чего он-лайн? Каким образом он-лайн?

Почему он-лайн?

Действительно, почему? Зачем маркетинговым исследователям Интернет? Еще совсем недавно, несколько лет назад, когда мы проводили первые проекты он-лайн, большинство клиентов относилось к Интернету крайне скептически. Бытовало мнение о том, что:

- в Интернет сообщество входит очень незначительная часть молодых людей, которая ни в коей мере не может репрезентировать российское население в целом;
- в сети общаются те, для кого в силу психологических проблем сложно общаться в реальной жизни, а также тем, кому делать нечего и, поэтому, они сутками сидят за компьютером;
- блоггеры и прочие сетевые фанаты всегда пишут не то, что думают и описывают события, которые никогда с ними не происходили.

Короче говоря, в Интернете можно изучать только ненормальных маргиналов, а нормальный, здоровый потребитель — герой нашего времени, дитя массовой культуры и мечта брэнд-менеджера — такими глупостями, как Интернет не занимается.

Прошло всего пару лет, и взгляды на Интернет сообщество сильно изменились. Да и сложно было бы им не измениться, если по данным на осень 2010 года полугодовая аудитория рунета (то есть те, кто пользуются Интернет хотя бы раз в полгода) составила 45 млн человек (и это только пользователи, старше 16 лет), вместе с подростками было бы на 5–6 млн больше. Это больше 30 % населения России. Если взять только горожан со средним и выше среднего доходом (а именно эта аудитория интересует маркетологов), то процент проникновения Интернета будем намного больше.

В общем, *size does matter*. Размер Интернет аудитории сегодня не позволяет говорить о том, что эта маргинальная группа. Скорее маргиналами являются те, кто сегодня не пользуется Интернетом (речь идет о больших городах и о более или менее состоятельных слоях населения, разумеется).

Еще одна причина, по которой на Интернет-аудиторию нужно обращать пристальное внимание, — огромный креативный потенциал этой среды. Интернет сообщества по всему миру давно подорвали монополию брэнд-менеджеров и рекламистов на производство имиджей. Сегодня каждый новый продукт, каждая более или менее заметная реклама немедленно становятся предметом обсуждения в сети. Причем, попав на страницы блогов и форумов, реклама начинает жить собственной жизнью, видоизменяется и порой приобретает совершенно не тот смысл, который в нее вкладывали создатели. Креативность и влияние Интернет сообществ просто не позволяют пренебрегать их мнением, особенно когда дело касается разработки новых продуктов или тестирования рекламы.

Наконец он-лайн открывает перед исследователями дополнительные возможности. Например, можно обсуждать определенную тему в течение достаточно долгого времени — несколько дней в отличие от нескольких часов на фокус-группе. К обсуждению можно привлекать людей, разбросанных географически. В он-лайн беседе респонденты имеют возможность немедленно проиллюстрировать свою точку зрения фотографией, видеозаписью или ссылкой на какой-то релевантный Интернет ресурс. Короче говоря, возможностей достаточно много, но много и ограничений, о них мы поговорим позднее.

Для чего он-лайн

С помощью он-лайн можно решать большинство исследовательских задач, но особенно эффективно данная группа методик работает:

- для изучения аудитории моложе 30 лет. В данном случае он-лайн формат более предпочтителен, потому что для этих людей Интернет является более привычной средой коммуникации. Весь их образ жизни, привычки и навыки общения сформированы таким образом, что в он-лайне они чувствуют себя более комфортно, чем в комнате для проведения фокус-групп, печатать на клавиатуре им удобнее, чем писать от руки (а надо сказать, что многие клиенты до сих пор предпочитают, чтобы респонденты заполняли дневники или лайф-буки в бумажном виде), стиль и этикет общения, принятый в рунете, более привычен и потому позволяет им высказываться более свободно и откровенно;
- для изучения трендов и поиска новых бизнес-ниш. Большинство тенденций, связанных с модой на различные новинки, изменением вкусов потребителей, покупательского поведения, да и образа жизни в целом зарождаются в Интернет сообществах и с их помощью же и очень быстро распространяются. Тренд, завоевавший популярность в Интернете, через какое-то время выплескивается в офф-лайн и дальше продолжает распространяться среди более консервативной части потребителей. Таким образом, для того чтобы засечь тренд на стадии его зарождения, уловить изменение предпочтений целевой аудитории и успеть воспользоваться возникшей возможностью, необходимо постоянно мониторить состояние Интернет сообществ;
- для создания новых продуктовых и рекламных концепций. Я уже упоминал о том, что Интернет является в настоящее время наиболее креативной средой, причем креативность присуща не столько самой среде (хотя и это тоже имеет место быть), сколько людям, которые в этой среде общаются, т. е. участникам Интернет сообществ. Разработка новых продуктовых концепций (NPD) — это вообще-то говоря, один из самых распространенных типов качественных исследований. Существует масса методик, которые, по сути, похожи и зачастую отличаются лишь названием *Concept Lab*, *Concept Development*, etc. Для участия в таких группах, как правило, специально приглашаются наиболее креативные респонденты. Здесь от респондента ждут не только и не столько откровений о его образе жизни и привычках, сколько новых идей. Получается, что респондент здесь, вроде бы и не совсем респондент, а своеобразный эксперт. Он-лайн методы дают широчайший простор для использования креативного потенциала потребителей. Причем особая прелесть Интернет среды в том, что она

позволяет очень быстро и точно определять тех, кто действительно способен генерить новые идеи.

Каким образом он-лайн?

В настоящее время практически каждая уважающая себя исследовательская компания имеет в своем арсенале методы для проведения качественных исследований он-лайн. Несмотря на то, что названия «брендируемых» исследовательских продуктов могут очень сильно отличаться, по сути, все методы можно разделить на 4 основных группы:

Internet observation/Internet Listening/Mystery chatting, etc.

Суть метода: В рунете идентифицируются наиболее влиятельные сообщества, посвященные категории, которую необходимо исследовать. Специально подготовленный менеджер проекта участвует в обсуждениях. Существует две стратегии участия: «новичок» и «опытный пользователь». Основная сложность здесь заключается именно в подготовке менеджера. В рунете вообще, а также в каждом сообществе в частности приняты свои правила обсуждения и особый сетевой этикет. Если его не соблюдать, никакой ценной информации получить не удастся.

В ходе *mystery chatting*, менеджер проекта идентифицирует наиболее влиятельных экспертов в категории, к чьему мнению Интернет-сообщество прислушивается, так называемых категориальных трендсеттеров. На заключительном этапе этих трендсеттеров приглашают на личное интервью. И здесь, принципиально важна техника проведения этого интервью. Чем более откровенно исследователь объясняет респонденту суть проекта, какую именно информацию заказчик хочет получить, какие именно бизнес-задачи с помощью этой информации решить, тем больше толку от такого интервью. Использование эксперта «в темную» в данном случае совершенно не подходит. Такая техника модерации получила название *co-creation*: исследователь, респондент, а в идеале и сам заказчик исследования вместе работают над проблемой.

Обычно такое исследование занимает четырех до пяти недель от момента начал *mystery chatting* до сдачи отчета.

Он-лайн сообщество

Суть метода. В отличие от *mystery chatting*, где мы наблюдаем жизнь других Интернет сообществ, в данном случае, мы сами такое сообщество и создаем. Участники сообщества рекрутируются он-лайн по различным критериям (пол, возраст, вовлеченность в категорию, пользование продуктом / маркой, степень авторитетности в глазах окружения, и т. д.). После этого мы создаем закрытое Интернет сообщество. На сегодняшний день существует достаточно много платформ, созданных специально для проведения качественных маркетинговых исследований, например 20/20, Visions Life, etc.

Особенность рекрута таких респондентов состоит в том, что мы с самого начала вводим их в курс дела, рассказываем о целях и задачах исследования, предельно подробно объясняем правила игры. Таким образом, можно сказать, что меняется статус участников исследования, из респондентов они превращаются в сотрудников, что в свою очередь очень сильно повышает их вовлеченность в процесс.

Далее модератор или несколько модераторов задают/ постят вопросы для обсуждения, как правило, 1 или 2 вопроса в день. Установочные вопросы носят общий характер, их цель — стимулировать дискуссию между членами сообщества. Например, модератор задает общий *life style* вопрос: «Как современный житель мегаполиса может вести здоровый образ жизни?». В процессе обсуждения, уточняется, какую роль в этом играет здоровое питание, какие именно продукты и бренды ассоциируются со здоровым питанием и т. д.

В общем, это похоже на то, как это происходит на обычной фокус-группе. Отличие в том, что процесс растянут во времени и респонденты физически друг с другом не общаются. Такой подход имеет ряд преимуществ. Во-первых, респонденты имеют возможность подумать, взвесить свои ответы. Во-вторых, у каждого участника есть возможность (и модератор, всячески поощряет их этой возможностью воспользоваться) проиллюстрировать свою мысль различными ссылками, цитатами, картинками и т. д. Наконец, в-третьих, сама непринужденная, естественная атмосфера дискуссии способствует более серьезному, глубокому отношению респон-

дентов к обсуждению. В результате, на выходе получаем больше информации и информация эта более качественная. В зависимости от цели исследования, категории, целевой аудитории, в сообщество приглашают от 12 до 50 человек. Срок работы такого сообщества составляет от недели до десяти дней.

Пролонгированное он-лайн сообщество

Часто в бизнес практике бывают такие ситуации, когда необходима целая серия исследовательских мероприятий, растянутая на несколько месяцев. Например, на первом этапе проходит изучение образа жизни ЦА, поиск бизнес ниш. На втором этапе создается новый продукт, или брэнд, а далее разрабатывается концепция его позиционирования. В этом случае имеет смысл поддерживать жизнь сообщества дольше, например, на протяжении нескольких месяцев. Участники сообщества постоянно находятся, что называется на «низком старте» и как только у клиента возникает необходимость подключить к процессу креативный потенциал потребителей, сообщество реанимируется и продолжает работу. На разных этапах с его участниками могут проводиться фокус-группы и глубинные интервью.

Он-лайн фокус-группа

По сути он-лайн группа очень похожа на обычную фокус-группу. Рекрутируют 6–8 участников соответствующим определенным критериям. Все участники получают доступ к платформе (их, опять же на сегодняшний день существует довольно много) и в определенный день и час выходят в Интернет.

Он-лайн формат облегчает рекрут респондентов, которые настолько заняты, что им намного удобнее поучаствовать в 2 часовой он-лайн группе, при этом находясь у себя дома или в офисе. Кроме того, он-лайн формат позволяет собрать респондентов, разбросанных географически. Все это, несомненно, работает в плюс.

Однако есть и ряд минусов. Во-первых, очень немногие платформы дают изображение приемлемого качества (зачастую либо картинка, либо звук сильно искажаются, что, естественно вызывает раздражение, как у респондентов, так и у модератора). Во-вторых, скорость и надежность Интернет каналов, до сих пор оставляют желать лучшего, особенно в региональных городах (сбои в работе Интернета также приводят к тому, что либо картинка, либо звук искажаются или «тормозят»). Что касается самого способа проведения он-лайн групп, то он в целом мало отличается от обычных групп и никаких дополнительных преимуществ ни в модерации, ни при анализе не дает.

Критика он-лайн методов

Конечно, он-лайн методы небезупречны. Предваряя критические замечания, я бы хотел остановиться на основных недостатках. Пожалуй, наиболее часто упоминаемым недостатком является отсутствие личного контакта между участниками он-лайн сообщества. Не видят они друг друга, не слышат интонации голоса, не могут наблюдать за невербальными реакциями. Именно поэтому, противники он-лайн утверждают, что в виртуальных сообществах отсутствует групповая динамика. Это, конечно, не так.

Участники он-лайн сообществ точно так же подчиняются законам группы. В сообществе, точно так же всегда выделяются свои лидеры и аутсайдеры, формируются коалиции и контркоалиции, а обсуждение может быть очень эмоциональным. Вообще, общение в он-лайн формате насыщено эмоциями, достаточно вспомнить, как распространены знакомства и флирт в Интернете. Так что групповая динамика существует и здесь. Примечательно, что для стимулирования конструктивной работы он-лайн группы модератору требуются иные качества, нежели для поддержания динамики в обычной группе. Часто бывает, что блестящий модератор реальных групп ничего не может сделать с группой виртуальной и наоборот.

Другой недостаток, на который часто ссылаются критики он-лайн методов, заключается в виртуальности, т. е. анонимности респондентов. В сети, как известно, люди идентифицируются не по именам, а по никам. Анонимность позволяет сетевым пользователям придумать себе любую личину, абсолютно не соответствующую его реальной личности. Критики он-лайн утверждают, что сетевые респонденты будут дурачить модератора, а тот не сможет их проверить.

Здесь, безусловно, есть, о чем подумать. Действительно в сети не принято раскрывать свои данные первому встречному, но не принято и скрывать их. Случаи, когда девушки пишут от имени юноши или, например, молодые люди, изображают из себя умудренных сединами профессионалов, достаточно редки. И, что характерно, такие мистификаторы поддаются вычислению. Каждый пользователь оставляет о себе массу информации в разных местах сети. Чаще всего достаточно бывает произвести поиск, сравнить информацию о человеке с разных сайтов и его настоящая идентичность становится понятной. Кроме того, немаловажный момент заключается в том, что респондент может скрыть свой возраст, иногда доход, но знание категории симулировать не возможно, мастерство как говорится, не пропьешь.

Что касается заведомой мистификации модератора, то такие случаи тоже очень редки. Человеку вообще не свойственно врать без особой нужды, это касается как сетевых пользователей, так и потребителей, не охваченных Интернет культурой. Если бы это было не так, то возможность проведения любых социологических исследований была бы поставлена под сомнение.

Заключение

Как видим, он-лайн методы в качественных исследованиях часто весьма полезны и помогают получить информацию, которую сложно или невозможно было бы получить при использовании традиционных исследовательских методик. Часто, но не всегда. Нельзя сказать, что он-лайн однозначно прогрессивнее, чем офф-лайн. Все зависит от конкретных исследовательских задач, от специфики целевой аудитории и т. д.

Краткий обзор он-лайн методов показывает, что они не только не исключают проведение традиционных групп и интервью, но и наоборот, работают намного эффективнее в сочетании с традиционными методами. Поэтому противопоставлять их просто бессмысленно.

Не нужно бояться использовать он-лайн методы в качественных исследованиях; но при этом не нужно идеализировать их и забывать о присущих им ограничениях.