

Проблемы и перспективы измерения лонгтейла

Беседовал Алексей Дерик

Об особенностях измерения аудитории нишевых и локальных каналов мы поговорили с Денисом Белослюдовым, директором департамента специальных проектов «Агентства 2».

Что сегодня отрасль медиаизмерений может предложить небольшим каналам?

В классическом подходе, принятом на рынке, для оценки аудитории используются пиплметры. Пиплметровые панели дают точные результаты в измерении телесмотра крупных, основных каналов, например входящих в эфирные мультиплексы. А для измерений так называемого «длинного хвоста» — большого списка каналов с невысокими рейтингами — границы доверительного интервала, действительно, достаточно широки. Для измерения этого лонгтейла мы стремимся использовать не только панельный метод.

Предложение малым каналам корректных измерений довольно ограничено и по иной причине. В нашей стране на предоставление данных о рейтингах каналов, используемых в закупках рекламы, Роскомнадзором была уполномочена единственная компания — «Медиаскоп». Они давно работают на рынке России, и их данные являются уже общепризнанной валютой медиаизмерений.

Закон, определивший такой порядок работы, ограничивает порядок использования для продажи и покупки рекламы иные измерения и рейтинги. Иными словами, рынок медиаизмерений находится под государственным присмотром, чтобы была только одна валюта и никаких различий. А этот измеритель исследует довольно небольшое число каналов в интересах рекламного рынка и только по городам более 100 тыс. жителей, что составляет лишь половину населения страны. Мы собираем данные по более 500 версиям различных каналов, и эти данные так же соответствуют



международным стандартам. Кстати, мы участвовали с нашей панелью в конкурсе на национального медиаизмерителя в 2016 году, потому что уверены в своих данных и методе. Выиграл, конечно же, «Медиаскоп».

Проблема измерения лонгтейла актуальна и для вас?

Да, поскольку мы также работаем с панельной выборкой, данные по малым каналам имеют погрешность измерений. Но осуществляя измерения аудитории «Триколор ТВ», мы работаем в условиях необходимости давать заказчику качественные данные по всем каналам, присутствующим в «Триколор ТВ». Такое положение обязывает, и у нас довольно хорошо проработано это направление. Наша панель расставлена в соответствии с международными стандартами GGTAM, определяющими правила составления выборки и ротации панели, с высокой точностью отражая аудиторию оператора. Как и «Медиаскоп», в своей работе мы опираемся на ежегодное установоч-

ное исследование, которое позволяет нам выяснить характеристики генеральной совокупности (абонентской базы оператора), чтобы своевременно обновлять параметры выборки.

Мы изначально не использовали отдельное устройство — пиплметр, а устанавливали специальное ПО на спутниковый ресивер участников панели, которое позволяет регистрировать просмотр отдельных членов домохозяйств. В отличие от технологии Audio Matching, такой подход позволяет точно фиксировать не только канал, но и его версию (если, к примеру, существуют самостоятельные SD- и HD-версии).

В пакетах «Триколор ТВ» есть интересный нишевый контент, а также региональные и национальные каналы. Для примера, у «Триколор ТВ» порядка 320 каналов. Если мы возьмем выборку даже в 6 тыс. человек, расставленную по квотам «Медиаскопа», там будет недостаточно данных, чтобы достоверно измерять все 320 каналов. Поэтому, в отличие от аккредитованного измерителя, у нас повышенные требования к панели — мы не ограничиваемся только городами 100 тыс.+, поскольку абоненты «Триколор ТВ» есть по всей России, и отбираем участников панели по 12 квотам.

На сегмент локальных и нишевых каналов приходится не такая уж большая доля рекламных денег. Нужны ли эти измерения?

Да, нужны. Данные очень востребованы. Анализ аудитории нужен не только для того, чтобы продавать рекламу. Каналам нужна эта информация для закупки контента и формирования привлекательной для своей аудитории программной сетки.

Например, еще один потребитель подобных данных — операторы. Они стараются добавить в свои предложения больше интересных качественных тематических каналов для узких аудиторий. А в «Триколор ТВ» на эти измерения даже опираются некоторые внутренние процессы. Поэтому нам всегда нужны все более точные цифры.

Теоретически панель можно бесконечно расширять, но любые пиплметровые исследования ограничены количеством домохозяйств, поэтому всегда будет хвост из каналов с показателями около нуля, погрешность оценки которых слишком велика.

🔗 Вы как-то решаете эту проблему?

Да, мы идем в сторону гибридных измерений.

У нас есть дополнительный проект, RPD, или Return path data — регистрация данных от ресиверов с обратной связью, подключенных к Интернету. С любого приемника мы можем получать логи телесмотра для дальнейшего анализа.

Приемников много, и они каждый день отправляют информацию. В результате у нас набирается большой массив данных. Имея определенные методы их обработки, мы сможем дополнять результаты панельных исследований, в том числе по малым каналам.

Особенность такой системы в том, что она самообучается по мере накопления данных и постоянно адаптируется к изменениям поведения абонентов. Мы не можем создать, например, пять жестких схем телесмотра (домохозяйка, пенсионер и т. п.) и применить их ко всем измерениям. Появляются интересные эфирные события (Олимпийские игры, парад на 9 мая и тому подобное или экстренные политические события, которые сложно предсказать), а также сезонные колебания предпочтений. Поэтому система постоянно сравнивает имеющиеся у нее данные с накопленной ранее статистикой и с данными с панели, чтобы фиксировать подобные отклонения, а в будущем и прогнозировать тренды.

Метод позволяет анализировать любое телесмотрение на любой платформе — мы сможем обработать аналогичные данные с любых устройств. И здесь мы не ограничены стационарными телевизорами. Это последний тренд в потреблении контента — мультиплатформенный гибкий нелинейный просмотр. С ростом объема обрабатываемых данных увеличивается точность анализа.

🔗 Если анализ обратной связи дает такие точные измерения, нельзя ли вообще отказаться от панели?

Нельзя. Ограничение RPD в том, что данные не имеют социально-демографических характеристик. Мы получаем обезличенную статистику: знаем, что ресивер с таким-то ID находится в определенном регионе и «смотрит» такие-то каналы. Мы не знаем ничего о владельце ресивера, а также о том, насколько репрезентативна выборка ресиверов, в данный момент имеющих подключение к Интернету. Эти данные в чистом виде неприменимы. Возможно, там 70% — мужчины, а 30% — женщины (фактически отражена аудитория, которая подключает ресиверы к Интернету, а вовсе не все абоненты «Триколор ТВ»).

Пиплметровая панель необходима, чтобы решить эту проблему репрезентативности — добавить к собранным данным социально-демографические характеристики, обеспечив своего рода золотой стандарт (типичные паттерны телесмотра определенных групп зрителей), с которым сравниваются результаты анализов телесмотра. Так мы можем понимать, кто смотрел контент на ресивере, о котором мы не знаем ничего, кроме региона, и оценивать размер домохозяйства. И чем больше у нас данных, тем точнее мы можем предсказывать профиль зрителя.

Поэтому панель будет поддерживаться и развиваться — мы планируем ее улучшать и уточнять, увеличивая количество пиплметров.

Опыт аналогичных гибридных измерений уже есть в мире у других компаний, например у Nielsen.

🔗 Можно ли методами работы с big data получать социально-демографические характеристики непосредственно из большого числа логов телесмотра, без панельных исследований?

У нас данных пока не настолько много, чтобы моделировать профили домохозяйств с низкой погрешностью, в отрыве от пиплметровой панели.

В целом на рынке, конечно, есть колоссальные данные по телесмотру — у операторов IPTV, спутникового телевидения и иного цифрового, где используются приставки с обратным каналом. Каждый владелец данных их анализирует, принимая на основании результатов какие-то внутренние решения. Но рынок в целом консервативен. Аналитика на основе анализа big data пока кажется непонятной и никак не проверяемой, не соот-

ветствующей никаким стандартам (ввиду их отсутствия). Поэтому пока не ясно, как эти данные могут быть использованы для рынка отдельно от другой аналитики.

Но измерения эволюционируют. Сначала были дневниковые измерения, потом электронные панели. Мы просто еще не пришли к тому, чтобы на основе только лишь big data принимать какие-то стратегические решения.

🔗 А можно ли собирать данные с устройств прямо с социально-демографическими характеристиками (например, спросив самих абонентов)?

В таких случаях необходимы квалифицированное согласие пользователя и регистрация телесмотра каждого в отдельности. А это и есть принцип нашего проекта. Наш проект абсолютно прозрачен — мы идем легальным способом, заключая с каждым панельистом договор. Существует закон о защите персональных данных в России, право на частную жизнь и прочее, и мы не собираемся уходить на «темную сторону», собирая данные без ведома наших абонентов.

Конечно, мы все уже привыкли, что разные устройства собирают о нас данные. Вопрос в том, как пользователи отнесутся к тому, что ресивер будет явным образом о чем-то спрашивать у абонентов оператора. На мой взгляд, рынок к этому пока не готов.

🔗 Реально ли собрать данные вообще со всех устройств оператора, чтобы получить более полную статистику?

Для получения данных необходим доступ к Интернету, а он есть не везде. Оснастить каждую приставку собственным модемом — дорого и не очень целесообразно, поскольку модем еще не гарантирует покрытие сотовой связью там, где этот ресивер установлен. Да и замена ресиверов у всех абонентов — мероприятие сложное. Остается лишь мотивировать абонента постепенно переходить на более новые устройства с возможностью подключения к Интернету с помощью новых интересных для него услуг. Так можно распределить труд по доставке данных до оператора на всю абонентскую базу. А с ростом числа абонентов, подключенных к Интернету, востребованность гибридных исследований будет увеличиваться, поскольку они позволят точнее измерять и большие, и малые каналы в комплексе. ■