ЭКОНОМИКА И МАТЕМАТИЧЕСКИЕ МЕТОДЫ, 2012, том 48, № 3, с. 79–90

МАТЕМАТИЧЕСКИЙ АНАЛИЗ ЭКОНОМИЧЕСКИХ МОДЕЛЕЙ

ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МОДЕЛИ УПРАВЛЕНИЯ АВТОМАТИЗИРОВАННЫМ СБОРОМ МАРКЕТИНГОВОЙ ИНФОРМАЦИИ В РЕЧЕВОМ ПРЕДСТАВЛЕНИИ

© 2012 г. **А.А. Саакян**

(Москва)

Рассматривается проблема автоматизированного сбора первичной маркетинговой информации при помощи инструментального средства, фиксирующего информацию в речевом представлении. Описывается новая экономическая ситуация, в которую попадает фирма, проводящая маркетинговые исследования при помощи этого средства. Разрабатываются экономико-математические модели оптимального поведения фирмы в новой ситуации.

Ключевые слова: экономико-математические модели, маркетинговые исследования, первичная информация, голосовой интерфейс, распознавание речи.

ВВЕДЕНИЕ

Увеличивающаяся сложность современной мировой экономики, рост рынков, бурное развитие новых информационных технологий ведут к росту динамичности ситуации на рынке и неустойчивости положения его участников, повышению степени неопределенности и риска при принятии маркетинговых решений на всех уровнях. В такой ситуации значительно увеличивается ценность информации, необходимой для обоснования маркетинговых решений, и, соответственно, важность процедур ее сбора и предоставления – маркетинговых исследований.

Большая часть расходов на маркетинговое исследование сопряжена с этапом сбора первичной информации, т.е. информации, получаемой непосредственно от объекта исследования: "…процесс сбора (первичной) информации является длительным, дорогим и трудоемким, связан с коммуникационными проблемами" (Божук, Ковалик, 2004). Современные средства сбора и фиксирования первичной информации практически не включают методы автоматизации и являются интуитивными и традиционными: их оптимальность (т.е. минимальная величина трудовых и стоимостных затрат) необоснована. Оптимальные автоматизированные средства сбора и фиксирования первичной информации способны значительно снизить величину затрат на проведение маркетингового исследования.

Опрос является наиболее часто применяемым методом сбора первичной информации. В работе (Кириченко, Саакян, 2010) доказано, что оптимальное средство автоматизированного сбора первичной информации в количественном опросе, проводящемся в полевых условиях, должно использовать речевое (голосовое) представление информации. Средство фиксирует получаемую от респондента информацию в речевом представлении, автоматически распознает и преобразует ее в машинное представление, которое далее вводится в автоматизированную маркетинговую информационную систему (МИС). В работе (Кириченко, Саакян, 2010) установлено, что такое средство может быть разработано в рамках актуальной парадигмы моделирования речевого сигнала на основе современных алгоритмов распознавания речи и представляет собой комплекс из диктофона для фиксирования информации и программного средства, функционирующего на настольном компьютере или ноутбуке, для автоматического распознавания зафиксированной информации и ввода ее в МИС.

При применении в количественном опросе традиционного, не применяющего автоматизацию средства сбора информация фиксируется на бумажных бланках, а затем вручную преобразуется в машинное представление и вводится в информационную систему. Ручной ввод зафиксированной информации в автоматизированную МИС сопряжен не только с трудовыми и стоимостными затратами, но и с неизбежными ошибками. Кроме того, при проведении исследования на обширной территории бланки с информацией необходимо пересылать в центральный офис, что требует значительного времени. В такой ситуации маркетологи вынуждены либо увеличить продолжительность исследования, либо ограничить обследуемую выборку потребителей по территориальному признаку, в результате чего оценки измеряемых величин могут оказаться смещенными.

Инструментальное средство, использующее речевое представление информации, не только сокращает затраты на сбор за счет автоматизации ввода информации в МИС, но и позволяет вводить информацию параллельно с ее фиксированием (т.е. интервьюированием). Автоматически преобразованная в машинное представление информация может обрабатываться, сохраняться или мгновенно и без затрат пересылаться по сети Интернет, например, в центральный офис фирмы. Кроме того, полнофункциональное применение автоматизированного средства, в отличие от традиционного, возможно и вне офиса, поскольку его аппаратная часть (диктофон и ноутбук) является портативной. Указанные свойства нового средства дают возможность:

– вычислять промежуточные оценки измеряемых величин в ходе сбора и при необходимости корректировать параметры исследования, т.е. осуществлять динамическое управление ходом исследования;

– с минимальной задержкой относительно момента фиксирования информации передавать ее посредством Интернета, например, в центральный офис фирмы, проводящей исследование, что позволяет фирме организовать сеть региональных филиалов для формирования территориально сбалансированной выборки респондентов;

– производить сбор информации в географически удаленных точках, командируя из центрального офиса группу интервьюеров и наделяя их не только функцией сбора информации, но и преобразования ее в цифровой вид посредством системы распознавания, функционирующей на ноутбуке.

Поскольку перечисленные возможности возникают только при автоматизированном сборе информации, фирма, применяющая рассматриваемое инструментальное средство, попадает в новую экономическую ситуацию. Эта ситуация характеризуется рядом параметров, управляя значениями которых фирма может минимизировать свои издержки. Проанализируем новую экономическую ситуацию, определим параметры, влияющие на размеры издержек, и сконструируем оптимизационные экономико-математические модели, необходимые для обоснованного выбора значений параметров.

РАЗРАБОТКА МОДЕЛЕЙ УПРАВЛЕНИЯ СБОРОМ ИНФОРМАЦИИ

Пусть выборка респондентов состоит из непересекающихся классов (подмножеств, кластеров, слоев). Принцип деления генеральной совокупности на подмножества и отбора в обследуемую выборку элементов каждого подмножества несуществен для нашей задачи, т.е. выборка может быть стратифицированной, групповой или квотной (Беляевский, 2004; Божук, Ковалик, 2004; Черчилль, 2000). Отметим, что в зависимости от характеристик конкретных оснований деления (классификационных признаков) идентифицировать респондента как члена определенного класса возможно либо до интервью, либо после проведения интервью, получения и обработки информации.

Рассмотрим случай, где принадлежность респондента к определенному классу известна до интервью. Пусть фирма, проводящая маркетинговые исследования, имеет сеть региональных филиалов (в качестве филиала может рассматриваться и фирма-партнер). При проведении маркетингового исследования необходимо опросить заданное число респондентов каждого класса. Пусть для каждого филиала известна оценка числа доступных для опроса респондентов каждого класса. Также известно число интервьюеров, работающих в филиале. Параметры, определяющие затраты на проведение исследования – стоимость контакта с респондентом и среднее число интервью в день (определяющееся длительностью интервью и затрачиваемым на поиск респондента временем), – зависят как от класса респондентов, так и от особенностей конкретного филиала. Поскольку автоматизированное средство обеспечивает оперативную бесплатную передачу собранной информации в центральный офис, никакие другие характеристики филиала, например его географическое положение, не влияют на величину затрат. Необходимо оптимальным образом разделить выборку респондентов между филиалами фирмы, т.е. для каждого филиала определить число опрашиваемых респондентов каждого класса, так чтобы в сумме получить заданное число респондентов (т.е. выполнить задание), не превысив при этом заданной длительности исследования и минимизировав суммарные затраты.

Формализуем приведенное словесное описание в виде экономико-математической модели, которую назовем "Филиал". Обозначим через K_j, $j = 1, ..., N^K$ число респондентов класса j, которое необходимо опросить в ходе исследования. Для каждого филиала i, $i = 1, ..., N^F$, и каждого класса j респондентов заданы величины $Q_{i,j}$, $G_{i,j}$ и $T_{i,j}$ – максимальное число доступных для интервью респондентов, стоимость одного интервью и время, необходимое на поиск респондента и проведение интервью соответственно. Обозначим через R_i, $i = 1, ..., N^F$ число интервьюеров в филиале i; T^{max} – максимально допустимую длительность этапа сбора информации; $x_{i,j}$ – искомые величины – число респондентов класса j, опрашиваемых в филиале i, $i = 1, ..., N^F$, $i = 1, ..., N^K$.

Используя введенные обозначения, запишем ограничения задачи. Число опрошенных респондентов каждого класса должно быть не меньше заданного:

$$\sum_{i=1}^{N^F} x_{i,j} \geq K_j \quad \forall j = 1, ..., N^K.$$

Число опрошенных в каждом филиале респондентов каждого класса не может превышать объем доступной в этом филиале выборки: $x_{i,j} \leq Q_{i,j} \ \ \forall i = 1, ..., N^F, \ \ \forall j = 1, ..., N^K$. В каждом филиале длительность процедуры опроса заданного числа респондентов не должна превышать предельной величины:

$$\sum_{j=1}^{N^K} x_{i,j} T_{i,j} \leq T^{max} R_i \quad \forall i = 1, ..., N^F.$$

Кроме того, искомые величины должны быть целыми неотрицательными:

$$x_{i,j} \in Z^+ \quad \forall i = 1, ..., N^F, \ \ \forall j = 1, ..., N^K, \text{ где } Z^+ = N \cup \{0\}.$$

Целевая функция затрат запишется в виде: $C = \sum_{i=1}^{N^F} \sum_{j=1}^{N^K} x_{i,j} G_{i,j}$.

Таким образом, **модель "Филиал"** имеет вид:

$$\sum_{i=1}^{N^F} x_{i,j} \geq K_j \quad \forall j = 1, ..., N^K;$$

$$x_{i,j} \leq Q_{i,j} \quad \forall i = 1, ..., N^F \quad \forall j = 1, ..., N^K;$$

$$\sum_{j=1}^{N^F} x_{i,j} T_{i,j} \leq T^{max} R_i \ \ \forall i = 1, ..., N^F;$$

$$x_{i,j} \in Z^+ \quad \forall i = 1, ..., N^F, \ \ \forall j = 1, ..., N^K;$$

$$C = \sum_{i=1}^{N^F} \sum_{j=1}^{N^K} x_{i,j} G_{i,j} \to \min.$$

Построенная модель относится к классу задач целочисленного программирования и может решаться любым известным методом, например методом ветвей и границ (Корбут, Финкельштейн, 1969; Мину, 1990; Юдин, 2010).

Рассмотрим теперь более сложную ситуацию. Пусть фирма, проводящая маркетинговые исследования, планирует разместить сеть филиалов, используя долгосрочный прогноз параметров исследований, которые будут ей заказаны. Пусть известно прогнозируемое число исследований, время начала, максимальная длительность и размер вознаграждения за каждое исследование, необходимый объем выборки респондентов каждого класса. Вознаграждение за проведение исследования может быть получено фирмой, только если опрос проведен в указанные сроки и число опрошенных респондентов каждого класса не меньше заданного. Пусть для каждого региона известна оценка числа доступных для опроса респондентов каждого класса. Возможность опросить определенное число респондентов в конкретном регионе зависит от наличия в этом регионе достаточного числа свободных, т.е. незанятых в других исследованиях, интервьюеров. Требуется определить оптимальное число интервьюеров для каждого региона, максимизирующее общую прибыль фирмы, учитывая существование фиксированных, т.е. не зависящих от числа проведенных исследований, затрат на открытие филиалов и содержание интервьюеров.

Формальную экономико-математическую модель этой задачи назовем "Сеть". Обозначим параметры исследования q ($q = 1, ..., N^Q$) через $\left(K_q, T_q^{start}, T_q^{max}, S_q\right)$, где $K_q = \{K_{q,j}\}_{j=1,...,N^K} = \left\{K_{q,1}, ..., K_{q,N^K}\right\}$ – параметры выборки; $K_{q,j}$ – число респондентов класса j, опрашиваемых в ходе исследования; T_q^{start}, T_q^{max} и S_q – время начала, максимальная длительность исследования и величина вознаграждения соответственно. По-прежнему для каждого филиала i ($i = 1, ..., N^F$) и каждого класса j респондентов ($q = 1, ..., N^K$) заданы величины $Q_{i,j}$, $G_{i,j}$ и $T_{i,j}$ – максимального числа доступных для интервью респондентов, стоимости одного интервью и времени, необходимого на поиск респондента и проведение интервью соответственно. Правую временную границу периода прогноза обозначим через T^{max}. Затраты на содержание одного интервьюера в филиале i в течение всего периода времени $[0; T^{max}]$обозначим через H_i; F_i – затраты на открытие филиала i.

Введем обозначения для переменных: $x_{q,i,j}$ – число респондентов класса j, опрашиваемых в филиале i в ходе исследования q; r_i – число интервьюеров в филиале i; y_q – булевы переменные, признак того, что исследование q будет производиться; z_i – булевы переменные, признак того, что филиал i будет открыт ($q = 1, ..., N^Q$, $i = 1, ..., N^F$, $j = 1, ..., N^K$). Искомыми являются величины r_i, y_q и $x_{q,i,j}$.

Пусть время t дискретно и измеряется, например, в днях. Введем в рассмотрение дискретные функции $B_q[t]$ ($q = 1, ..., N^Q$), равные единице для тех значений t, в которых исследование q активно (т.е. уже началось и еще не завершилось), и равные нулю для остальных значений t:

$$B_q[t] = \begin{cases} 1, & t \in \left[T_q^{start};\ T_q^{start} + T_q^{max}\right], \\ 0, & t \notin \left[T_q^{start};\ T_q^{start} + T_q^{max}\right]. \end{cases}$$

Обозначим через $r_{q,i}[t]$ число интервьюеров филиала i, проводящих опрос в рамках исследования q в момент времени t. Очевидно, должно выполняться условие:

$$\sum_{q=1}^{N^Q} r_{q,i}[t] \le r_i \quad \forall i = 1, ..., N^F, \quad \forall t = 0, ..., T^{max}.$$

Так же как и в модели "Филиал", число опрошенных в каждом филиале респондентов каждого класса не может превышать объем доступной в этом филиале выборки:

$$x_{q,i,j} \le Q_{i,j} \quad \forall q = 1, ..., N^Q, \quad \forall i = 1, ..., N^F, \quad \forall j = 1, ..., N^K.$$

Это ограничение может быть усилено требованием участия одного респондента не более чем в одном исследовании. В этом случае последнее семейство неравенств заменяется на

$$\sum_{q=1}^{N^Q} x_{q,i,j} \le Q_{i,j} \quad \forall i = 1, ..., N^F, \quad \forall j = 1, ..., N^K.$$

Рассмотрим теперь ограничения, связанные с проведением конкретного исследования. Если исследование q проводится, т.е. $y_q = 1$, то аналогично модели "Филиал" число опрошенных для этого исследования респондентов каждого класса должно быть не меньше заданного:

$$\sum_{i=1}^{N^F} x_{q,i,j} \geq K_{i,j} \quad \forall q = 1, ..., N^Q, \quad \forall j = 1, ..., N^K;$$

интервьюеры могут работать в интересах этого исследования только в те дни, когда оно активно:

$$r_{q,i}[t] \geq 0, \text{ если } B_q[t] = 1;$$

$$r_{q,i}[t] = 0, \text{ если } B_q[t] = 0;$$

при этом интервьюеры должны успеть опросить заданное число респондентов:

$$\sum_{j=1}^{N^K} T_{i,j} x_{q,i,j} \leq \sum_t r_{q,i}[t] \quad \forall q = 1, ..., N^Q, \quad \forall i = 1, ..., N^F.$$

Если исследование q не проводится, т.е. $y_q = 0$, то для этого исследования число опрошенных респондентов каждого класса должно быть равно нулю:

$$x_{q,i,j} = 0 \quad \forall q = 1, ..., N^Q, \quad \forall i = 1, ..., N^F, \quad \forall j = 1, ..., N^K,$$

а число работающих интервьюеров также равно нулю:

$$r_{q,i}[t] = 0 \quad \forall q = 1, ..., N^Q, \quad \forall i = 1, ..., N^F, \quad \forall t = 0, ..., T^{max}.$$

Объединение ограничений, соответствующих разным значениям y_q, является обобщением ограничений на случай произвольного значения y_q, которое может быть равным нулю или единице, и запишется в виде:

$$My_q \geq \sum_{i=1}^{N^F} x_{q,i,j} \geq K_{q,i} y_q \quad \forall q = 1, ..., N^Q, \quad \forall j = 1, ..., N^K;$$

$$r_{q,i}[t] \leq MB_q[t] y_q \quad \forall q = 1, ..., N^Q, \quad \forall i = 1, ..., N^F, \quad \forall t = 0, ..., T^{max};$$

$$\sum_{j=1}^{N^K} T_{i,j} x_{q,i,j} \leq \sum_t r_{q,i}[t] \quad \forall q = 1, ..., N^Q, \quad \forall i = 1, ..., N^F,$$

где M – число, гарантированно большее максимально возможных значений величин $\sum_{i=1}^{N^F} x_{q,i,j}$ и $r_{q,i}[t]$.

Аналогично конструируются ограничения, связывающие значение признака открытия филиала с числом работающих в нем интервьюеров. Если филиал i будет открыт, то $z_i = 1$ и $r_i \geq 1$. В противном случае $z_i = 0$ и $r_i = 0$. Такое множество допустимых значений описывается системой неравенств:

$$\begin{cases} 0 \leq z_i \leq 1; \\ z_i \leq r_i; \\ Mz_i \geq r_i \quad \forall i = 1, ..., N^F. \end{cases}$$

Кроме того, все искомые величины должны быть целыми неотрицательными:

$$x_{q,i,j} \in Z^+, \ r_{q,i}[t] \in Z^+, r_i \in Z^+, \ y_q \in Z^+, \ y_q \leq 1, \ z_i \in Z^+, \ z_i \leq 1,$$

$$\forall q = 1, ..., N^Q, \quad \forall i = 1, ..., N^F, \ \forall j = 1, ..., N^K, \ \forall t = 0, ..., T^{max}.$$

Целевая функция равняется суммарному вознаграждению за все проведенные исследования, из которого вычтены суммарные затраты на проведение исследований, открытие филиалов и содержание интервьюеров:

$$C=\sum_{q=1}^{N^Q} y_q S_q-\sum_{q=1}^{N^Q}\sum_{i=1}^{N^F}\sum_{j=1}^{N^K} x_{q,i,j} G_{i,j}-\sum_{i=1}^{N^F}(z_i F_i+r_i H_i).$$

Таким образом, построенная **модель "Сеть"**, которая так же, как и модель "Филиал", имеет вид задачи целочисленного программирования, записывается в виде:

$$\sum_{q=1}^{N^Q} r_{q,i}[t]\le r_i,\ \ i=1,\dots,N^F,\ \ \forall t=0,\dots,T^{max};$$

а) $\sum_{q=1}^{N^Q} x_{q,i,j}\le Q_{i,j},\ \ \forall i=1,\dots,N^F,\ \ \forall j=1,\dots,N^K$ – если каждый респондент участвует не более чем в одном опросе;

б) $x_{q,i,j}\le Q_{i,j}\ \ \forall q=1,\dots,N^Q,\ \ \forall i=1,\dots,N^F,\ \ \forall j=1,\dots,N^K$ – если респондент участвует в произвольном числе опросов;

$$My_q\ge\sum_{i=1}^{N^F} x_{q,i,j}\ge K_{q,j}y_q\ \ \ \forall q=1,\dots,N^Q,\ \ \forall j=1,\dots,N^K;$$

$$r_{q,i}[t]\le MB_q[t]y_q\ \ \ \forall q=1,\dots,N^Q,\ \ \forall i=1,\dots,N^F,\ \ \forall t=0,\dots,T^{max};$$

$$z_i\le r_i\ \ \ \forall i=1,\dots,N^F;$$

$$Mz_i\ge r_i\ \ \ \forall i=1,\dots,N^F;$$

$$\sum_{j=1}^{N^K} T_{i,j}\,x_{q,i,t}\le\sum_t r_{q,i}[t]\ \ \ \forall q=1,\dots,N^Q,\ \ \forall i=1,\dots,N^F;$$

$$x_{q,i,j}\in Z^+,\ r_{q,i}[t]\in Z^+,\ r_i\in Z^+,\ y_q\in Z^+,\ y_q\le 1,\ z_i\in Z^+,\ z_i\le 1;$$

$$\forall q=1,\dots,N^Q,\ \forall i=1,\dots,N^F,\ \forall j=1,\dots,N^K,\ \forall t=0,\dots,T^{max};$$

$$C=\sum_{q=1}^{N^Q} y_q S_q-\sum_{q=1}^{N^Q}\sum_{i=1}^{N^F}\sum_{j=1}^{N^K} x_{q,i,j} G_{i,j}-\sum_{i=1}^{N^F}(z_i F_i+r_i H_i)\to\max.$$

Предположим теперь, что фирма, проводящая маркетинговые исследования, отказывается от содержания сети региональных филиалов, предпочитая просто командировать своих сотрудников-интервьюеров в другие города для проведения одного исследования, после чего они возвращаются в центральный офис. Пусть известно общее число интервьюеров фирмы, объем и структура выборки респондентов каждого города, время в пути и стоимость командирования одного сотрудника в этот город. Необходимо для каждого города определить число командируемых интервьюеров и опрашиваемых респондентов каждого класса, так чтобы в сумме получить заданное число респондентов, не превысив при этом заданной длительности исследования и минимизировав суммарные затраты.

Формальную экономико-математическую модель этой задачи назовем "Выезд". Так же как в модели "Филиал", обозначим через K_j, $j=1,\dots,N^K$, число респондентов класса j, которое необходимо опросить в ходе исследования. Для каждого города i ($i=1,\dots,N^F$) и каждого класса j респондентов ($j=1,\dots,N^K$) заданы величины $Q_{i,j}$, $G_{i,j}$ и $T_{i,j}$ – максимального числа доступных для интервью респондентов, стоимости одного интервью и времени, необходимого на поиск респондента и проведение интервью, соответственно. Обозначим через R общее число интервьюеров фирмы; T^{max} – максимально допустимую длительность этапа сбора информации; H_i – затраты на командирование одного интервьюера в город i; D_i – время в пути до этого города; $x_{i,j}$ – число респондентов класса j, опрашиваемых в городе i; r_i – число интервьюеров, командируемых в городе i; $i = 1, ..., N^F$, $j = 1, ..., N^K$. Искомыми являются величины $x_{i,j}$ и r_i.

Ограничение длительности этапа сбора информации запишется в виде:

$$\sum_{j=1}^{N^K} x_{i,j} T_{i,j} \le (T^{max} - D_i) r_i \quad \forall i = 1, ..., N^F,$$

а целевая функция затрат –

$$C = \sum_{i=1}^{N^F} \sum_{j=1}^{N^K} x_{i,j} G_{i,j} + \sum_{i=1}^{N^F} r_i H_i.$$

В остальном формальная **модель "Выезд"** совпадает с моделью "Филиал":

$$\sum_{i=1}^{N^F} x_{i,j} \ge K_j, \quad \forall j = 1, ..., N^K,$$

$$x_{i,j} \le Q_{i,j}, \quad \forall i = 1, ..., N^F, \ \forall j = 1, ..., N^K,$$

$$\sum_{j=1}^{N^K} x_{i,j} T_{i,j} \le (T^{max} - D_i) r_i, \quad \forall i = 1, ..., N^F,$$

$$\sum_{i=1}^{N^F} r_i \le R,$$

$$x_{i,j} \in Z^+, \quad \forall i = 1, ..., N^F, \ \forall j = 1, ..., N^K,$$

$$C = \sum_{i=1}^{N^F} \sum_{j=1}^{N^K} x_{i,j} G_{i,j} + \sum_{i=1}^{N^F} r_i H_i \to \min.$$

Рассмотрим теперь второй случай деления генеральной совокупности респондентов на классы: принадлежность респондента к определенному классу становится известна только после проведения интервью, или даже сам принцип деления на классы корректируется в ходе сбора информации после получения и обработки очередной ее подвыборки.

Предположим, что фирме, проводящей маркетинговые исследования, необходимо опросить заданное число респондентов каждого класса в условиях, когда идентификация респондента как члена определенного класса возможна только после обработки полученной от него информации. В условиях подобной неопределенности общее число респондентов, опрошенных в ходе исследования, может превысить заданное вначале (до сбора информации). Затраты на интервьюирование респондентов сверх необходимого числа фирме не компенсируются, т.е. являются убытками. Пусть контакт с респондентами возможен в нескольких местах (точках), причем для каждого задано априорное распределение вероятностей принадлежности респондента классу. Такими точками могут быть, например, магазины одной торговой сети: структура выборки посетителей даже в очень похожих по ассортименту, размеру и оформлению магазинов может сильно различаться, поскольку зависит от их территориального положения, соседства с учебными заведениями, станциями метро, шоссе и т.п. Предположим, в начале каждого рабочего дня фирма может направлять своих интервьюеров в одну или несколько точек, выбираемых из общего пула. Необходимо, исходя из известных априорных распределений вероятностей принадлежности респондентов классам, разместить интервьюеров по точкам так, чтобы минимизировать математическое ожидание величины убытков от "сверхплановых" интервью.

Формальную экономико-математическую модель этой задачи назовем "Точка". Обозначим через K_j, $j = 1, ..., N^K$, число респондентов класса j, которое осталось опросить для завершения исследования, т.е. невыполненный пока остаток задания. Предполагается, что оптимальное планирование при помощи модели "Точка" осуществляется на срок, равный одному дню, т.е. оптимальное распределение интервьюеров корректируется ежедневно на основе скорректированных величин K_j. Для каждой точки i, $i = 1, ..., N^F$ заданы величины: $\bar{P}_i = (P_{i,1}, ..., P_{i,N^K})$ – априорное распределение вероятностей принадлежности респондента классу; R_i – вместимость; S_i – скорость.

Величины $P_{i,j}$ являются вероятностями того, что очередной респондент, опрошенный в точке i, принадлежит классу j. Вместимостью называется максимальное число интервьюеров, которые могут одновременно проводить опрос в этой точке. Скорость точки является оценкой среднего числа респондентов, которых может опросить один интервьюер за период планирования (один день).

Обозначим: R – число интервьюеров фирмы; T – время, оставшееся до завершения этапа сбора информации; r_i – искомое число интервьюеров, командируемых в точку i, $i = 1, ..., N^F$. Очевидно, что должны выполняться условия:

$$\sum_{i=1}^{N^F} r_i \le R, \quad r_i \le R_i \quad \forall i = 1, ..., N^F.$$

Примем в качестве допущения, что фактическое (апостериорное) число $m_{i,j}$ респондентов класса j, опрошенных в точке i в течение одного дня, линейно зависит от числа интервьюеров в этой точке: $m_{i,j} = c_{i,j} r_i$, где $c_{i,j}$ – число респондентов класса j, опрошенных в точке i в течение одного дня одним интервьюером, $C_{i,j}$ – случайная величина с известными математическим ожиданием $\bar{c}_{i,j} = S_i P_{i,j}$ и дисперсией $\sigma^2_{i,j}$. Математическое ожидание $\bar{m}_{i,j}$ величины $m_{i,j}$ равно $\bar{m}_{i,j} = \bar{c}_{i,j} r_i = S_i P_{i,j} r_i$.

Введем также обозначения

$$m_j = \sum_{i=1}^{N^F} m_{i,j}, \quad m = \sum_{j=1}^{N^K} m_{i,}, \quad \bar{m}_j = \sum_{i=1}^{N^F} \bar{m}_{i,j}, \quad K = \sum_{j=1}^{N^K} K_j.$$

Если по условиям задачи для результирующей выборки респондентов требуется обеспечить заданную пропорцию классов независимо от объема выборки, следует минимизировать математическое ожидание суммы превышения фактического числа респондентов каждого класса над заданным. Заданное число респондентов некоторого класса – это фактическое число респондентов (просуммированное по всем классам), умноженное на заданный долевой коэффициент этого класса. Тогда целевая функция запишется в виде:

$$C = M\left[\sum_{j=1}^{N^K} (m_j - \hat{m}_j)\right] \to \min,$$

где $\hat{m}_j = \min\left(m_j, [mK_j/K]\right)$, [ ] – целая часть числа.

Указанная комбинация требований к выборке респондентов может представлять исключительно теоретический интерес, поскольку необходимость в выборке с такими свойствами на практике не встречается. Практический же интерес представляет вторая постановка задачи, требующая обеспечить выполнение плана $\{K_1, ..., K_{N^K}\}$ при заданном числе респондентов каждого класса, минимизировав число "холостых" контактов. Иначе говоря, вероятность того, что фактическое число респондентов каждого класса окажется не меньше заданной величины K_j, должна быть больше или равна заданной величине α, а минимизируемая целевая функция убытков является апостериорным суммарным (по всем классам) числом респондентов:

$$P(Tm_j \ge K_j) \ge \alpha \quad \forall j = 1, ..., N^K; \quad C = M\left[T\sum_{j=1}^{N^K} m_j\right] \to \min.$$

Полностью **модель "Точка"** записывается в виде:

$$\sum_{i=1}^{N^F} r_i \le R;$$

$$r_i \le R_i, \quad i = 1, ..., N^F;$$

$$r_i \in Z^+;$$

$$P\left(T\sum_{i=1}^{N^F} c_{i,j}\, r_i \ge K_j\right) \ge \alpha, \quad j = 1, ..., N^K;$$

$$C = TM\left[\sum_{j=1}^{N^K}\sum_{i=1}^{N^F} c_{i,j}\, r_i\right] = T\sum_{j=1}^{N^K}\sum_{i=1}^{N^F} S_i P_{i,j}\, r_i \to \min.$$

Тем самым получена линейная задача стохастического программирования в М-постановке (Юдин, 2010). Если можно принять условие независимости величин $c_{i,j}$ и соответствия их распределения нормальному закону, то построенная стохастическая задача сводится к эквивалентной детерминированной задаче выпуклого программирования с линейной целевой функцией и квадратичными ограничениями (Юдин, 2010). Вероятностные ограничения в этом случае перепишутся в виде:

$$T\sum_{i=1}^{N^F} S_i P_{i,j}\, r_i - \Phi^{-1}(\alpha)\left(\sum_{i=1}^{N^F} \sigma_{i,j}^2\, r_i^2\right) \ge K_j, \quad j = 1, ..., N^K.$$

Будем считать, что $\alpha \ge 0{,}5$. В этом случае $\Phi^{-1}(\alpha) \ge 0$ и $T\sum_{i=1}^{N^F} S_i P_{i,j} \ge K_j$, а область допустимых решений, ограничиваемая приведенными условиями, выпукла. Полностью эквивалентная детерминированная задача запишется в виде:

$$\sum_{i=1}^{N^F} r_i \le R;$$

$$r_i \le R_i, \quad i = 1, ..., N^F;$$

$$r_i \in Z^+;$$

$$T\sum_{i=1}^{N^F} S_i P_{i,j}\, r_i - \Phi^{-1}(\alpha)\left(\sum_{i=1}^{N^F} \sigma_{i,j}^2\, r_i^2\right) \ge K_j, \quad j = 1, ..., N^K;$$

$$C = \sum_{j=1}^{N^K}\sum_{i=1}^{N^F} S_i P_{i,j}\, r_i \to \min.$$

Для решения этой задачи можно применять любой из подходящих методов, например один из методов возможных направлений (Базара, Шетти, 1982; Зуховицкий, Авдеева, 1967).

Построенную модель "Точка" можно комбинировать с моделями вычисления оценки априорного распределения вероятностей принадлежности респондента к классу (для каждой точки), где используются как априорные знания, так и выборочные данные. Целесообразность корректирования априорного распределения вероятностей непосредственно в ходе сбора информации должна определяться для каждого конкретного случая отдельно, исходя из прогнозируемой достоверности таких оценок.

ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА ЭФФЕКТИВНОСТИ ОПТИМАЛЬНОГО ПЛАНИРОВАНИЯ

Для экспериментальной оценки качества разработанных экономико-математических моделей организации сбора маркетинговой информации и измерения величины снижения издержек было произведено оптимальное распределение заданий по сбору информации между региональными филиалами фирмы, проводящей маркетинговое исследование методом полевого количественного опроса-интервью при помощи автоматизированного средства сбора. Для эксперимента было выбрано маркетинговое агентство "Точка роста". Исходные данные для применения оптимизационной модели были получены специалистами маркетингового агентства методом экспертных оценок.

Совокупность респондентов была разбита на классы в соответствии с наблюдаемым признаком – местом встречи интервьюера и респондента. Число респондентов каждого класса,

Таблица 1. Классы респондентов

№	Класс респондентов	Место встречи интервьюера с респондентом	Число респондентов в выборке
1	Покупатели продуктов для домашнего хозяйства	Сетевые супермаркеты	1200
2	Покупатели одежды	Специализированные магазины одежды	600
3	Студенты	Помещения и прилегающие территории вузов	900
4	Туристы	Центральные площади города, памятники архитектуры	450
5	Посетители музеев	Выход из музея	600
6	Посетители кинотеатров	Выход из кинотеатра	600

Таблица 2. Характеристики региональных филиалов

№	Город	Число интервьюеров в филиале
1	Москва (центральный офис)	50
2	Санкт-Петербург	20
3	Тверь	8
4	Пермь	12
5	Екатеринбург	10
6	Новосибирск	20
7	Красноярск	15
8	Владивосток	10

Таблица 3. Экспертные оценки элементов матриц $\|Q_{i,j}\|$, $\|G_{i,j}\|$ и $\|T_{i,j}\|$

Город	Покупатели продуктов	Покупатели одежды	Студенты	Туристы	Посетители музеев	Посетители кинотеатров
Москва	10000/50/25	10000/50/20	10000/30/18	10000/50/18	10000/50/18	10000/50/20
Санкт-Петербург	1000/44/30	1000/44/23	5000/30/17	10000/44/15	5000/36/18	1000/44/20
Тверь	200/36/40	100/36/40	700/24/15	50/36/25	50/36/35	200/36/30
Пермь	300/40/40	300/40/25	600/36/20	1500/40/20	700/40/25	400/40/20
Екатеринбург	400/34/40	200/34/40	400/30/15	100/34/30	50/34/35	200/34/30
Новосибирск	500/36/35	350/36/30	5000/20/15	300/36/25	300/36/35	250/36/20
Красноярск	400/40/35	350/40/30	300/40/20	100/40/25	200/40/35	300/40/20
Владивосток	300/44/40	300/44/30	400/44/20	350/44/30	100/44/35	300/44/30

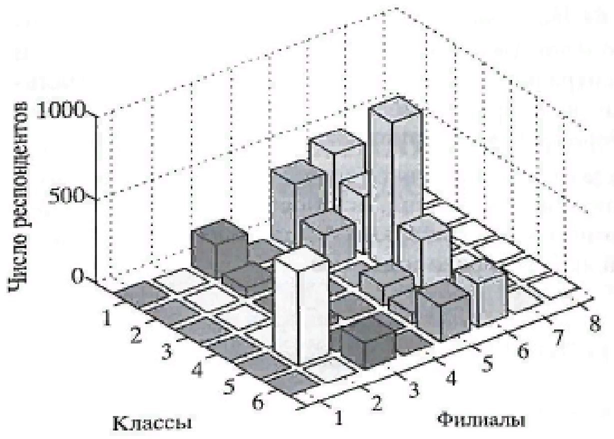

Рис. 1. Вычисленное оптимальное распределение выборки респондентов

Рис. 2. Базовое распределение выборки респондентов

необходимое для проведения исследования, приведено в табл. 1. Наличие априорной информации о классе респондента позволяет применить для оптимизации модель "Филиал". Список региональных филиалов вместе с числом работающих в каждом из них интервьюеров представлен в табл. 2. Экспертные оценки матриц: $\|Q_{i,j}\|$ – максимальное число доступных в каждом городе респондентов каждого класса, $\|G_{i,j}\|$ – стоимость одного интервью в рублях, $\|T_{i,j}\|$ – длительность одного интервью в минутах – представлены в табл. 3. Каждое поле табл. 3 имеет формат QQQQ/GG/TT, где QQQQ – число респондентов, GG – стоимость интервью, TT – длительность интервью. Длительность этапа сбора информации была установлена равной десяти дням при восьмичасовом рабочем дне интервьюеров.

Согласно определению модели "Филиал", задача оптимального распределения выборки респондентов между региональными филиалами является задачей целочисленного линейного программирования. Решение этой задачи приведено в числовом виде в табл. 4, а также в графическом виде на рис. 1.

На рис. 2 представлено базовое распределение, реализуемое в случае отсутствия филиалов или возможности их задействовать. Базовое распределение, очевидно, предписывает производить сбор информации исключительно силами центрального офиса.

Для получения оценки относительной эффективности оптимального распределения заданий по сбору информации между филиалами необходимо сравнить достигающуюся в этом случае

Таблица 4. Вычисленное оптимальное распределение выборки респондентов

Город	Покупатели продуктов	Покупатели одежды	Студенты	Туристы	Посетители музеев	Посетители кинотеатров
Москва	0	0	0	0	0	0
Санкт-Петербург	0	0	0	0	550	0
Тверь	200	50	0	50	0	150
Пермь	0	0	0	0	0	0
Екатеринбург	400	200	0	100	50	200
Новосибирск	500	350	900	300	0	250
Красноярск	100	0	0	0	0	0
Владивосток	0	0	0	0	0	0

величину затрат с аналогичной величиной для базового распределения. Величина затрат составила 140 700 руб. для оптимального планирования (в случае использования филиалов) и 199 500 руб. для случая использования только центрального офиса. Таким образом, оптимальный план применения автоматизированного средства сбора первичной информации уменьшил затраты на 29,5 % за счет распределения задач сбора между региональными филиалами.

Разработанные экономико-математические модели организации сбора маркетинговой информации позволяют фирме, проводящей маркетинговые исследования, осуществлять оптимальное управление ресурсами – региональными филиалами и интервьюерами – для снижения затрат и формирования территориально сбалансированной выборки респондентов.

СПИСОК ЛИТЕРАТУРЫ

Базара М., Шетти К. (1982): Нелинейное программирование. Теория и алгоритмы. М.: Мир.

Беляевский И.К. (2004): Маркетинговое исследование: информация, анализ, прогноз. М.: Финансы и статистика.

Божук С.Г., Ковалик Л.Н. (2004): Маркетинговые исследования. СПб.: Питер.

Зуховицкий С.И., Авдеева Л.И. (1967): Линейное и выпуклое программирование. М.: Эдиториал УРСС.

Кириченко А.А., Саакян А.А. (2010): Сбор первичной маркетинговой информации при помощи голосового интерфейса // *Маркетинг*. № 4. С. 47–57.

Корбут А.А., Финкельштейн Ю.Ю. (1969): Дискретное программирование. М.: Наука.

Мину М. (1990): Математическое программирование. Теория и алгоритмы. М.: Наука.

Черчилль Г.А. (2000): Маркетинговые исследования. СПб.: Питер.

Юдин Д.Б. (2010): Задачи и методы стохастического программирования. М.: Эдиториал УРСС.

Поступила в редакцию
23.06.2011 г.

Economical-Mathematical Models of Management of Automated Collecting the Marketing Information in the Speech Representation

A.A. Saakian

This article considers the problem of primary marketing information automated collecting. The new tool that collects information using its speech representation is used for the automation of this task. The new economic circumstances that arise when company uses that tool for marketing researches described. The mathematical models of new economic circumstances developed.

Keywords: economical mathematical model; marketing research; primary information; voice interface; speech recognition.