

СОВРЕМЕННЫЕ МЕТОДЫ И ИНСТРУМЕНТЫ СЕГМЕНТИРОВАНИЯ ПОТРЕБИТЕЛЬСКИХ РЫНКОВ

Мартышенко Н.С.¹, Грачева В.В.¹

¹*Владивостокский государственный университет экономики и сервиса, Владивосток, Россия (690014, Владивосток, Гоголя, 41), e-mail: Natalya.martyshenko@vvsu.ru*

Многообразие задач сегментирования и условий формирования рынков породило множество методов сегментирования. Целью работы является составить научный путеводитель в пространстве современных методов и инструментов сегментирования потребительских рынков. Среди методов сегментирования наиболее мощным инструментом являются методы многомерной классификации данных. В работе рассмотрены различные подходы построения алгоритмов многомерной классификации. При исследовании социально-экономических систем подавляющее большинство явлений не поддается прямому измерению (умственные способности, личностные качества, толерантность, компетентность, мобильность, политические убеждения и т.д.). В таких случаях рекомендуется использовать латентные модели или латентный структурный анализ. В практике сегментирования часто приходится иметь дело с качественными данными. Обработка качественных данных строится на принципах типологий, которые воплощаются в компьютерных технологиях. Использование комбинации количественных и качественных методов часто является наилучшим решением проблемы сегментирования рынка.

Ключевые слова: сегментирование рынка, многомерный анализ данных, латентный структурный анализ, анализ качественных данных

MODERN METHODS AND TOOLS OF SEGMENTATION CONSUMER MARKETS

Martyshenko N.S.¹, Gracheva V.V.¹

¹*Vladivostok state university economy and service, Vladivostok, Russia (690014, Vladivostok, street Gogol, 41), e-mail: sergey.martishenko@vvsu.ru*

The variety of problems of segmentation and conditions of formation of the markets have generated set of methods of segmentation. The work purpose is to make the scientific guidebook in space of modern methods and tools of segmentation of the consumer markets. Among segmentation methods the most powerful tool are methods of multidimensional classification of the data. In work various approaches of construction of algorithms of multidimensional classification are considered. At research of social and economic systems the overwhelming majority of the phenomena does not give in to direct measurement (mental faculties, personal qualities, tolerance, competence, mobility, political convictions etc.) . In such cases it is recommended to use latent models or the latent structural analysis. In segmentation practice often it is necessary to deal with the qualitative data. Processing of the qualitative data is under construction on principles typologization which are embodied in computer technologies. Use of a combination of quantitative and qualitative methods often is the best decision of a problem of market segmentation.

Keywords: market segmentation, multidimensional analysis of data, latent structural analysis, analysis of qualitative data

Основным стимулом развития любого рынка являются покупатели. Множество покупателей неоднородно. Они отличаются по вкусам, желаниям и мотивам совершения покупок. Это явление породило такой важнейший инструмент исследования рынков, как сегментирование рынка, целью которого является выработка гибкой политики взаимодействия с различными группами потребителей.

В настоящее время сложилось множество подходов к сегментированию рынков, среди которых трудно ориентироваться практическим работникам. Поэтому в работе была предпринята попытка составить научный путеводитель в пространстве современных методов и инструментов сегментирования потребительских рынков.

Сегмент рынка (market segment) – совокупность потребителей, характеризующихся однотипной реакцией на предлагаемые продукты и другие элементы комплекса маркетинга.

Сегментирование рынка представляет собой формальную процедуру, основанную на применении методов многомерного статистического анализа к результатам маркетинговых исследований.

При сегментировании потребительского рынка сначала используют объективные переменные сегментирования (социально-экономические, демографические, географические), а затем субъективные переменные сегментирования (психографические, поведенческие, личностные). Из полного перечня признаков сегментирования обычно выбирают один или несколько наиболее важных. Большое число переменных сегментирования может привести к чрезмерному дроблению сегмента.

Выделяют два подхода к сегментированию рынков: «a priori» и «post hoc».

В рамках первого подхода считается, что предварительно известны признаки сегментирования, емкость сегментов, их количество, характеристики, карта интересов. То есть предполагается, что сегментные группы потребителей продуктов в данном методе уже сформированы («a priori»).

В рамках второго подхода («post hoc»), предполагается неопределенность признаков сегментирования и сущности самих сегментов. Необходимо произвести поиск признаков сегментирования с последующим выбором и описанием сегментов.

Многообразие задач сегментирования и условий формирования рынков породили множество методов сегментирования. В настоящее время в практике исследования рынков получили широкое распространение следующие методы сегментирования:

- метод многомерной классификации;
- метод корреляционного сегментирования – «K-сегментирование»;
- методы архетиповой сегментации;
- метод сегментации по выгодам;
- метод построения сетки сегментации;
- метод группировок;
- метод функциональных карт;
- метод сегментации на основе матриц Абеля;
- метод сегментации потребителей по степени их лояльности;
- метод сегментации по выгодам.

Среди методов сегментирования, которые появились достаточно недавно, можно назвать следующие методы:

- метод коллаборативной фильтрации;

- метод латентных моделей;
- метод гибкого (flexible) сегментирования;
- метод компонентного (componential) сегментирования.

Среди методов сегментирования наиболее мощным инструментом являются методы многомерной классификации данных. Применение метода многомерной классификации называют кластерным анализом.

Кластерный анализ это совокупность методов, позволяющих классифицировать многомерные наблюдения, каждое из которых описывается набором исходных переменных $X_1, X_2, \dots, X_j, \dots, X_m$. Целью кластерного анализа является образование групп сходных между собой объектов, которые принято называть кластерами.

Задачу классификации нужно отличать от задачи группировки. Задача группировки состоит в том, что данные разбиваются сначала по уровням одного признака, затем по уровням другого признака и т.д. В отличие от задачи группировки в кластерном анализе формирование групп объектов (классов) производится с учетом всех группировочных признаков одновременно.

Для решения задачи классификации необходимо ввести понятие сходства объектов по наблюдаемым признакам. В каждый класс должны попасть объекты, обладающие определенной степенью сходства.

Для количественной оценки сходства вводится понятие метрики. Сходство между объектами будет определяться в зависимости от расстояния в выбранном метрическом пространстве. Если объект, описываемый m признаками представить точкой в m -мерном пространстве, то сходство объектов друг с другом будет определяться как расстояние в данном метрическом пространстве.

В кластерном анализе используется большое разнообразие способов измерения расстояний (метрик). Примером одной из наиболее распространенные метрик сходства является евклидово расстояние:

$$d_{ij} = \sqrt{\sum_{r=1}^m (x_{ir} - x_{jr})^2} \quad (1)$$

где d_{ij} – расстояние между i -м и j -м объектами;

x_{ir}, x_{jr} – значение r -го признака соответственно у i -го и j -го объекта;

$i = \overline{1, n}$, где n – объем выборки;

$j = \overline{1, m}$, где m – количество признаков.

Для решения задачи классификации необходимо рассматривать расстояния между каждой парой объектов. Расстояния между парами объектов сводятся в матрицу сходства. Это симметричная матрица. По диагонали матрицы располагаются нулевые значения.

Оценивать сходство объектов с помощью мер расстояния удобно при использовании числовых признаков. Но часто встречаются признаки, измеренные в других шкалах (например, в ранговой, или, вообще, в номинальной). В этом случае все признаки, используемые для классификации, приводятся к представлению в бинарном (двоичном) коде. Предположим, что такое преобразование было выполнено. То есть каждый объект описывается вектором $\overline{X}_i = (x_{i1}, x_{i1}, \dots, x_{im})$ $i = \overline{1, n}$, каждая из компонент которого принимает значения 0 или 1.

Для измерения сходства i -го и j -го объектов введем следующие обозначения частот:

$n_{ij}^{(1,1)}$ – число совпадающих единичных признаков у обеих пар объектов (пар (1,1));

$n_{ij}^{(0,0)}$ – число совпадающих нулевых признаков у обеих пар объектов (пар (0,0));

$n_{ij}^{(1,0)}$ – число совпадающих единичных признаков у i -го и нулевых признаков у j -го объектов (пар (1,0));

$n_{ij}^{(0,1)}$ – число совпадающих нулевых признаков у i -го и единичных признаков у j -го объектов (пар (0,1));

$n_i^{(1)}, n_j^{(1)}$ – число единичных признаков у i -го и единичных признаков у j -го объектов соответственно;

$n_i^{(0)}, n_j^{(0)}$ – число нулевых признаков у i -го и нулевых признаков у j -го объектов соответственно;

$p_{ij} = n_{ij}^{(1,1)} + n_{ij}^{(0,0)}$ – общее число совпадающих признаков;

$q_{ij} = n_{ij}^{(0,1)} + n_{ij}^{(1,0)}$ – общее число несовпадающих признаков;

m – общее число признаков, по которым осуществляется сравнение.

В качестве примеров мер подобия в бинарной шкале измерения можно привести коэффициент Рао (2) и коэффициент Хаммана (3):

$$S_{ij} = \frac{n_{ij}^{(1,1)}}{m} \quad (0 \leq S_{ij} \leq 1) \quad (2)$$

$$S_{ij} = \frac{p_{ij} - q_{ij}}{m} \quad (\text{и} \text{ } \ddot{\partial} \text{ } p_{ij} = q_{ij} \quad S_{ij} = 0) \quad (3)$$

Другой важной величиной в кластерном анализе является расстояние между целыми группами объектов. Приведем примеры наиболее распространенных расстояний и мер близости, характеризующих взаимное расположение отдельных групп объектов. Пусть ω_t – t -я группа (класс, кластер) объектов, n_t – число объектов, образующих группу ω_t , вектор μ_t – среднее арифметическое объектов, входящих в группу ω_t (другими словами μ_t – «центр тяжести» t -й группы), а $\psi_{tf}(\omega_t, \omega_f)$ – расстояние между группами ω_t и ω_f .

Наиболее распространенными методами определения расстояния между кластерами ω_t и ω_f являются: метод «ближайшего соседа» (4), метод «дальнего соседа» (5), метод оценки расстояния между центрами тяжести (6).

$$\psi_{tf}^{\min}(\omega_t, \omega_f) = \min_{x_i \in \omega_t, x_j \in \omega_f} d(x_i, x_j) \quad (4)$$

$$\psi_{tf}^{\max}(\omega_t, \omega_f) = \max_{x_i \in \omega_t, x_j \in \omega_f} d(x_i, x_j) \quad (5)$$

$$\psi_{tf}^*(\omega_t, \omega_f) = d(\mu_t, \mu_f) \quad (6)$$

Среди всех методов классификации самым распространенными являются иерархические агломеративные методы. Основная идея этих методов состоит в том, что на первом шаге каждый объект выборки рассматривается как отдельный кластер. Иерархическая процедура состоит в пошаговом объединении наиболее близких классов. Близость классов оценивается по матрице расстояний или матрице сходства. На первом шаге матрица сходства имеет размерность $n \times n$. На следующем шаге при объединении двух классов матрица сходства пересчитывается. Размерность матрицы сокращается на единицу и становится $[(n-1) \times (n-1)]$. Процесс завершается за $n-1$ шагов, когда все объекты будут объединены в один класс.

Процесс объединения объектов можно изобразить в виде графа-дерева (дендрограммы). На дендрограмме указываются номера объединяемых объектов и расстояния, при которых произошли объединения. При выделении классов руководствуются скачками метрики сходства на дендрограмме.

Большое распространение в практике экономического анализа получил метод многомерной классификации, который известен под названием «метод k -средних». Он был предложен Мак-Куином. Этот метод классификации относится к группе итеративных методов классификации. Существует множество модификаций этого метода. Рассмотрим алгоритм классификации в его первоначальном виде.

Пусть имеется n объектов (наблюдений), каждый из которых характеризуется m признаками $X_1, X_2, \dots, X_j, \dots, X_m$. Необходимо разбить наблюдения на заданное число классов – k .

Шаг ноль. Из n точек исследуемой совокупности случайным образом отбирается k точек. Эти точки принимаются как центры классов.

Итерация. Множество точек разбивается на k классов по минимуму расстояния до центров классов. Для расчета расстояния можно использовать любую метрику. Чаще всего используется евклидово расстояние. Производится пересчет центров классов, как центров тяжести точек, присоединенных к классам.

Проверка. Если центры классов при выполнении очередной итерации не изменились, то процесс классификации завершается, иначе переходим к пункту «итерация».

Существует большая группа алгоритмов многомерной классификации, основанная на применении теории графов. Представителем этой группы является алгоритм «Краб» [2]. На сходных принципах организована группа алгоритмов FOREL.

Рассмотрим принципы работы алгоритма «Краб». Работа алгоритма начинается с нахождения пары точек с минимальным расстоянием между ними. Эти точки соединяются ребром графа. Затем соединяются следующие самые близкие точки, из числа не присоединенных, к уже построенной части графа. Эта процедура повторяется до тех пор, пока все точки не окажутся соединенными ребрами графа.

Такой граф не будет иметь петель, и суммарная длина всех его ребер будет минимальной. Граф, обладающий такими свойствами, называется кратчайшим незамкнутым путем (КНП). Для разбиения множества точек графа на два таксона разрывают самое длинное ребро. Далее процесс повторяется до получения адекватной структуры классов. Алгоритм «Краб» имеет множество модификаций. При дальнейшем развитии алгоритма вводится понятие λ – расстояния и критерии качества разбиения.

При исследовании социально-экономических систем подавляющее большинство явлений не поддается прямому измерению (умственные способности, личностные качества, толерантность, компетентность, мобильность, политические убеждения и т.д.), что привело к появлению нового понятия «латентные» переменные. Для исследования латентных переменных используются латентные модели или латентный структурный анализ (ЛСА). Это достаточно широкий класс моделей, которые нашли свое применение в различных областях, в том числе, при решении задач сегментирования.

Анализ латентной структуры (от лат. *Latentis* – скрытый, невидимый) – это статистический анализ эмпирических данных, позволяющий по ответу респондентов на некоторое множество вопросов выявить их распределение по некоторому скрытому

(латентному) признаку. Этот признак нельзя измерить непосредственно, но использованное множество вопросов позволяет зафиксировать различные его проявления. Метод предложен известным американским социологом Полом Лазарсфельдом.

Суть модели, предложенной Лазарсфельдом, сводилась к следующему. Предполагается, что существует некоторая латентная переменная, которая объясняет внешнее поведение респондентов. Это поведение можно объяснить, анализируя ответы каждого человека на определенные дихотомические вопросы анкеты. Латентная переменная номинальна, число ее значений заранее известно исследователю. Объясняющая способность латентной переменной обуславливается тем, что именно она служит причиной наличия связи между наблюдаемыми переменными [4].

В основе классического латентного структурного анализа, лежит фундаментальная аксиома Лазарсфельда локальной независимости: при фиксации значения латентной переменной связи между наблюдаемыми переменными исчезают.

Подсовокупности респондентов с одинаковыми значениями латентной переменной образуют латентные классы.

Различают несколько типов латентных моделей: модели для непрерывной латентной переменной; модели для дискретной латентной переменной; модели для дихотомических признаков.

Рассмотрим общую модель латентного структурного анализа для дискретной латентной переменной, поскольку в анкетных опросах чаще имеют дело именно с такими признаками.

Будем считать, что латентная переменная является прерывной, а респонденты расположены в дискретных точках этой переменной – классах.

Пусть имеется m классов латентной переменной α и n дихотомических вопросов. Обозначим:

p_i – вероятность ответить положительно на i -й вопрос (доля тех, кто ответил положительно на i -й вопрос)

p_{ij} – вероятность ответить положительно на i -й и j -й вопросы одновременно.

$p_{i,j,k}$ – вероятность ответить положительно на i, j, k -й вопросы.

p_{ijk} – вероятность ответить отрицательно на i -й вопрос и положительно на j -й и k -й вопросы

v^m – доля респондентов, попавших в класс m .

p_i^m – вероятность ответить положительно на i -й вопрос при условии попадания в класс m .

В соответствии с формулой полной вероятности и аксиомой локальной независимости, составим основные расчетные уравнения:

$$\sum_{\alpha=1}^m v^{\alpha} = 1, \quad (7)$$

$$p_{\sigma} = \sum_{\alpha=1}^m p_{\sigma}^{\alpha} v^{\alpha}, \quad (8)$$

где $\alpha=1, \dots, m$, а σ – набор индексов.

Приведем пример системы уравнений для $m=2, n=2$

$$\begin{cases} v^1 + v^2 = 1 \\ p_1^1 v^1 + p_1^2 v^2 = p_1 \\ p_2^1 v^1 + p_2^2 v^2 = p_2 \\ p_1^1 p_2^1 v^1 + p_1^2 p_2^2 v^2 = p_{12} \end{cases}, \quad (9)$$

где p_1, p_2 – известны, $p_1^1, p_1^2, p_2^1, p_2^2, v^1, v^2$ – не известны.

В общем виде модели латентных классов число уравнений – 2^n , а число неизвестных параметров – $m(n+1)$. Очевидно, что для того, чтобы система уравнений имела решения, необходимо, чтобы число неизвестных не превосходило число уравнений, т.е. $2^n \geq m(n+1)$.

Преимуществом латентного структурного анализа по сравнению с кластерным анализом является то, что этот метод позволяет включать в анализ переменные, измеренные в разных шкалах.

Одним из подходов к оценке латентных переменных является использование модели Раша.

Измерения с помощью модели Раша – это процесс трансформации исходных данных тестового вида в интервальную шкалу натуральных логарифмов. Для вычисления латентных переменных в модели Раша вводится понятие «логит».

Логит – это условная единица, легко переводимая в любую другую шкалу. В виду того, что шкала Раша интервальная, это позволяет использовать большое количество различных процедур статистического анализа. Помимо этого, в интервальной шкале нулевая точка отсчета (0) не зафиксирована, поэтому показатели в логитах переводятся в другую систему показателей, например, баллов при помощи линейных преобразований. А за точку отсчета в системе логитов наиболее уместно принять среднее значение показателей наблюдаемых переменных.

Первоначально модели Раша были использованы для анализа знаний при тестировании. В настоящее время область применения моделей распространилась и на другие объекты. Например, аппарат моделей Раша может быть применен для структурного анализа данных анкет, включающих вопросы, характеризующие толерантность к различным социально-экономическим процессам. Толерантность может выражаться в выборе ответов из списка:

- сильное согласие (конечно, да);
- слабое согласие (скорее да, чем нет);
- слабое несогласие (скорее нет, чем да); сильное несогласие (конечно, нет).

Основным уравнением для политомической модели Раша (переменные варьируются более чем на двух уровнях) является уравнение:

$$P\{x_{ji} = x\} = \frac{\exp \sum_{l=0}^x [\beta_j - (\tau_l + \delta_i)]}{\sum_{k=0}^m \exp \sum_{l=0}^x [\beta_n - (\tau_l + \delta_i)]} \quad (10)$$

где x – градация индикаторной переменной (варьируются от 0);

x_{ji} – оценка j -го респондентом для i -го индикатора (пункт опросника);

$P\{x_{ji} = x\}$ – вероятность выбора j -ым респондентом градации x для i -го индикатора;

β_j – месторасположение j -го респондента на шкале «эффективность социально-экономической толерантности» (измеряется в логитах);

δ_i – месторасположение i -го индикатора на той же шкале;

τ_l – относительный оценка l -ой градации;

m_j – индексная переменная, которая последовательно принимает все варианты ответов на индикативные вопросы.

Эта модель позволяет измерить в одной и той же шкале (в логитах) уровень толерантности и информативность пунктов опросника [7].

Далее рассмотрим несколько, относительно, новых методов сегментирования.

Один из таких методов представлен в работе [8]. Этот метод основан на применении теории графов.

Рассматриваемая математическая модель сегментации рынка услуг базируется на 3-дольном 3-однородном гиперграфе $G = (V, E) = (V_1, V_2, V_3, E)$. Вершины первой доли, т.е. $v \in V_1$, взаимно однозначно соответствуют элементам множества предоставляемых предприятием или группой предприятий услуг. Каждая вершина второй доли $v \in V_2$ однозначно соответствует некоторому элементу из множества потребителей,

классифицированного по демографическим признакам. Вершины третьей доли $v \in V_3$ взаимно однозначно соответствуют элементам множества потребителей, классифицированного по признаку «уровень доходов»: низкий, ниже среднего, средний, выше среднего, высокий, элит-класс. Для построения множества ребер $E = \{e\}$ рассматриваются всевозможные тройки вершин v_1, v_2, v_3 – такие, что $v_1 \in V_1, v_2 \in V_2, v_3 \in V_3$. Всякую такую тройку называют допустимой, если услуга V_1 может быть приемлема для потребителей данного уровня доходов V_3 и для данной демографической категории V_2 . Множество всех ребер $E = \{e\}$ определяется как множество всех допустимых троек $\square e = (v_1, v_2, v_3), v_i \in V_i, i = \overline{1,3}$.

В гиперграфе $G = (V, E) = (V_1, V_2, V_3, E)$ допустимым решением рассматриваемой задачи сегментации рынка услуг является всякий такой его подгиперграф $\square x(V_x, E_x), V_x \subseteq V, E_x \subseteq E$, в котором каждая компонента связности представляет собой простую звезду с центром в вершине $v \in V_1$. Через $X = X(G) = \{x\}$ обозначим множество всех допустимых решений задачи покрытия гиперграфа G звездами. Например, одно из таких решений представлено на рис. 1.

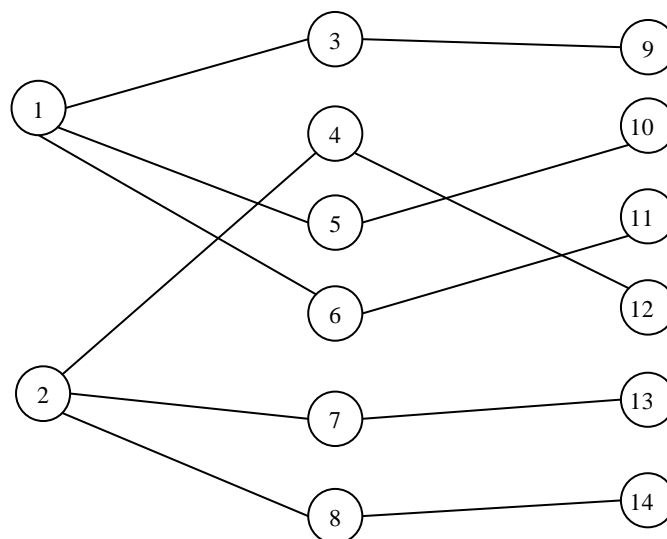


Рис. 1. Допустимое покрытие графа звездами

Для численной оценки качества допустимых решений каждому ребру $e \in E$ гиперграфа $G = (V, E)$ приписываются три веса $w_\nu(e), \nu = \overline{1,3}$, которые представляют собой экспертные оценки. В качестве весов могут быть: мощность услуги в данном позиционировании, ожидаемая устойчивость данного позиционирования, и другие. Качество допустимых решений этой задачи оценивается с помощью векторной целевой функции (11).

$$F_\nu(x) = \sum_{e \in E_x} w_\nu(e) \rightarrow \max, \nu = \overline{1,3} \quad (11)$$

В последнее время все большее распространение находит метод сегментирования основанный на коллаборативной фильтрации [10].

В общем виде алгоритм коллаборативной фильтрации можно описать так: существует множество пользователей $\{u_1, u_2, \dots, u_n\}$ и множество объектов $\{b_1, b_2, \dots, b_m\}$. Каждый пользователь имеет список оцененных им объектов. Оценки могут принадлежать разным шкалам от 1 до 10, от 1 до 5 и т.д., а также разным типам шкал: порядковой или относительной. Если пользователь u_i желает получить рекомендацию (или прогноз своей оценки на неоцененный им объект), то по известным оценкам устанавливаются ближайшие по предпочтениям (или по оценкам на одни и те же объекты) пользователи к множеству u_i . Далее алгоритм выдает рекомендации пользователю u_i (или рассчитывает прогнозную оценку на объект), исходя из оценок ближайших к u_i пользователям по их предпочтениям.

Собственная концепция сегментирования на основе стоимости была предложена Институтом оценки бизнеса компании IBM (рис. 2) [9].

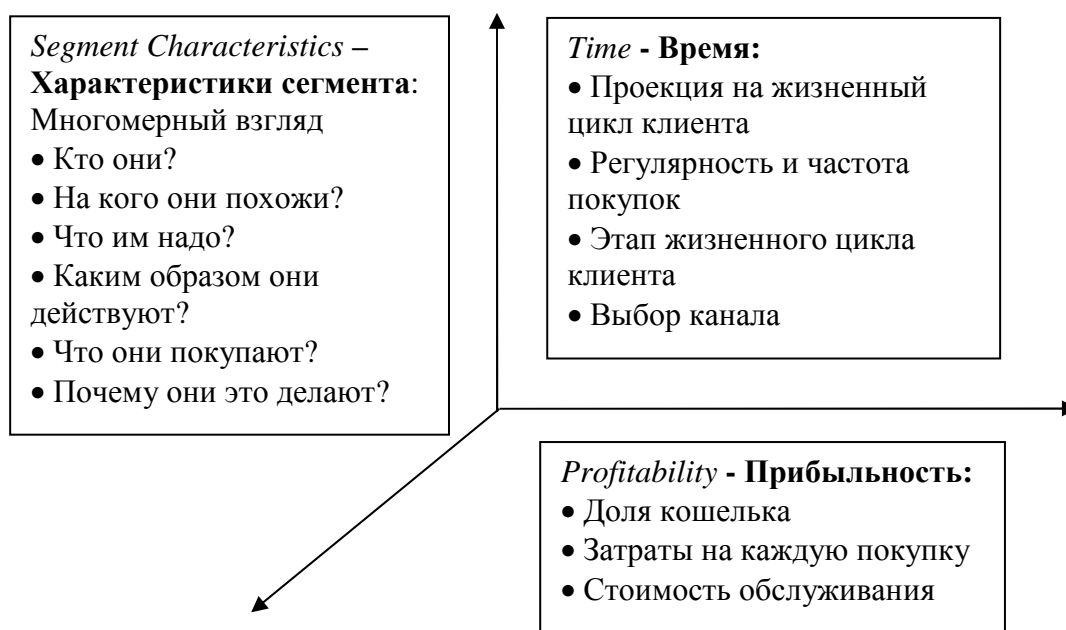


Рис. 2. Принцип сегментирования, использованный институтом оценки бизнеса компании IBM

В туристском бизнесе широкое распространение нашли методы сегментации рынка по психографическим признакам. Сегментация рынка по психографическим признакам – это процесс разделения всех покупателей рынка на однородные группы по таким критериям, как: ценности, убеждения, мотивация к покупке товара и тип личности.

Сегментирование по психографическим признакам основывается на теории типологического анализа.

Типологический анализ – метаметодика анализа данных, совокупность методов изучения социального феномена, позволяющих выделить социально значимые, внутренне

однородные, качественно отличные друг от друга группы эмпирических объектов, характеризующиеся типобразующими признаками, природа которых различна, и интерпретируемые как носители различных типов существования феномена [11].

Типологический анализ, в основном, базируется на анализе качественных данных. По мнению В.А. Ядова, качественные методы позволяют глубже понять изучаемое явление и предложить множественную интерпретацию [12].

В мировой литературе утвердилось понятие «качественные методы исследования», которые иногда называют «мягкими» в отличие от «количественных» и «жестких». В определенном смысле можно говорить также о неформализованных или слабо формализованных подходах в сравнении с жестко формализованными.

Основанием типологии служит совокупность суждений (утверждений) о близости (схожести, похожести) объектов, носителей информации об изучаемых социальных феноменах (явлениях, процессах).

Предметом типологии является совокупность основных характеристик социального феномена, ответственная за отнесение эмпирических объектов к однотипной группе.

Рассмотрим некоторые термины, которые используются в типологическом анализе.

Тип (анг. type) – вид, форма существования социальных феноменов в науке или в повседневной жизнедеятельности людей. Тип – сущность, знание о которой всегда относительно. Имеет три условных значения в смысле – типовой, типологический, типический.

Типологический – в типологическом анализе: особенный, общий, объединяющий.

Типизация (анг. typization) – конструирование людьми социальной реальности на основе придания окружающим ярлыков, спонтанная классификация.

Типологизация (анг. typologization) – процедура систематизации знаний об изучаемых феноменах либо для введения (задания) типов, либо для поиска знаний о типах. Типологизация служит для конструирования типов.

Типология (анг. typology) – Совокупность типов, результат их конструирования. Способ конструирования типов.

Типообразующий признак — характеристика, свойство социальных феноменов, на основе которых либо конструируются типы, либо формулируются гипотезы об их существовании. Типообразующий признак – это концептуальная переменная.

Существует несколько подходов к решению проблемы неструктурированных или неформализованных данных. В саму природу нечисловой информации заложена возможность использования для ее обобщения и структуризации типологический анализ. «При всей уникальности действующего индивида большая часть его индивидуальных

смыслов типична, т. е. обладает общностью с другими людьми» [1]. Одной из основных задач является разработка алгоритма построения типологий таким образом, чтобы преодолеть субъективность исследователя, не упустив при этом важную информацию. Алгоритмы предполагают применение сжатия и структурирования информации так, чтобы она сохранила свойства исследуемого объекта.

Можно выделить три подхода построения типологий:

- концепция типологических операций А. Бартона и П.Ф. Лазарсфельда;
- анализ структуры У. Герхардта;
- типологический анализ У. Кукартца.

Первый подход основан на использовании «типологических операций»: редукции, субструкции и трансформации. Через определение признаков и степеней их выраженности строится пространство свойств, которые лежат в основе типологии. Используя графическую или табличную формы представления данных, определяются все возможные комбинации и все потенциальные типы. Все типы связываются на одном пространстве. Комбинации признаков могут быть сокращены. Субструкция выявляет само пространство признаков, лежащее в основе типологии, которое может трансформироваться при интерпретации сконструированных типов. Несмотря на то, что данный подход был предложен для конструирования типов в исследованиях количественной стратегии, он имеет центральное значение для обобщения нечисловой информации.

Второй и третий подходы используют те же основные типологические операции, пытаясь при этом преодолеть его главный недостаток: определение критериев отбора признаков для анализа данных. В основе второго подхода лежат «идеальные типы», которые служат базисом для анализа информации, полученной в ходе исследования. На первом этапе проводится сравнение случаев через их реконструкцию, чтобы выявить их особенности. Это привносит в исследование прозрачность процесса обобщения и его результатов. На втором этапе исследуемые случаи группируются с помощью их сопоставления. Эти приемы в целом соответствуют «концепции типологических операций» и позволяют узнать все потенциально возможные комбинации признаков. На последнем этапе выявляются и объясняются смысловые связи внутри и между полученными группами. Для этой цели был разработан анализ структуры и процесса, состоящий из двух шагов обобщения. Главными недостатками данного подхода являются трудность абстрагироваться от субъективных представлений исследователя при обобщении данных и отсутствие алгоритма проверки сконструированных типов.

Типологический анализ в данном подходе имеет ряд особенностей по сравнению с другими методами обобщения качественных данных. В процессе работы абстрагируются от

каждого отдельного исследуемого случая и получают типичное событие, как результат упорядоченных фаз секвенции. «Структурная герменевтика», наоборот, понимает материал в его единичности, неотделенности от каждого конкретного исследуемого случая. Типологический анализ находится в точке пересечения между индивидуальной историей и общепринятой. Во втором подходе исследуемые случаи сохраняются по возможности в своей целостности, в третьем подходе при анализе отдельных случаев и их сравнении используется тематическое обобщение.

Третье направление в полной мере нельзя считать отдельным подходом. Это «инструмент, построенный для целей выражения методологических взглядов М. Вебера» [3]. При разработке инструментов типологического анализа (программных средств) предпринимается попытка соединить в исследовании различные способы типологизации нечисловой информации с учетом их достоинств и недостатков.

Известные социологи Н. Филдинг и Р. Ли в предметной области инструментальных средств анализа качественных данных предложили использовать специальный термин «компьютерный, но ассистируемый анализ качественных данных» (Computer Assisted Qualitative Data Analysis, CAQDAS). Современный компьютерно-ассистируемый анализ качественных данных является методологической исследовательской областью, объединяющей ученых многих стран. Ассистируемый анализ представлен множеством компьютерных пакетов, в том числе: Atlas.ti, MAXQDA, NVivo, xSight, Qualrus, Ethnograph и др. Эти пакеты представляют собой класс компьютерных программ, которые включают в свою архитектуру специальные структуры, называемые функциями кодирования и реконструирования качественных, нечисловых данных (coding and retrieval functions). Функции кодирования и реконструирования данных (ФКР) представляют собой компьютерный инструмент (tool), используемый в человеко-машинном режиме и ассистирующий пользователю при изучении данных, представленных в так называемых нечисловых форматах. В основе ассистирования лежит аппарат аналитических переобозначений, называемых кодами, введение и связывание которых между собой осуществляется самим пользователем. Более подробный анализ методологических разработок компьютерного инструментария анализа качественных данных представлен в работе [3].

Отмечая широкие возможности зарубежных компьютерных инструментальных средств анализа качественных данных нельзя не отметить, что они не нашли своего распространения не только у отечественных ученых, занимающихся проблемами сегментирования рынка, но и у отечественных социологов. Эти средства имеют несколько

другую направленность и больше предназначены для решения гуманитарных и лингвистических задач.

В работах [5, 6] представлена методология анализа качественных данных, которые собираются с целью исследования рынков и, в частности, туристского рынка. Методология включает достаточно простые в применении компьютерные программные инструменты, которые могут быть использованы в рамках компьютерной среды EXCEL, которая является наиболее распространенной среди отечественных исследователей рынков. Предложенная методология основывается на представлении данных в форме «термов».

Термом называется символьное выражение: $t(X_1, X_2, \dots, X_n)$, где t – имя терма, называемая функтор или «функциональная буква», а X_1, X_2, \dots, X_n – термы, структурированные или простейшие. Для формального описания термов в работе было введено новое понятие – составной признак.

Предложенную методологию анализа данных, используемых при сегментировании, необходимо рассматривать не столько как учение о методах, сколько как учение о взаимодействии методов между собой на разных классах исследовательских практик анализа данных. Структурно эта методология включает приемы и методы сбора и измерения информации, а также математические методы.

Использование комбинации количественных и качественных методов часто является наилучшим решением проблемы сегментирования рынка. Различные методы дополняют и контролируют друг друга, ограничения одного метода уравниваются ограничениями другого. Такие свойства называют комплементарностью и триангуляцией.

Комплементарными называют несходные или даже противоположные теории, концепции, модели и точки зрения, отражающие различные взгляды на действительность.

Триангуляция – это возможность использования несколько источников информации. В анализе рынка можно выделить несколько типов триангуляции: триангуляция данных; триангуляция исследователей; триангуляция методов; триангуляция теорий.

Список литературы

1. Готлиб, А.С. Введение в социологическое исследование. Качественный и количественный подходы. Методология. Исследовательские практики: учебное пособие – М: Флинта: МПСИ, 2005. – 384 с.
2. Загоруйко Н.Г. Прикладные методы анализа данных и знаний – Новосибирск: Изд-во Института математики, 1999. – 270 с.

3. Каныгин Г.В. Инструментальные средства и методологические принципы анализа качественных данных // Социология: методология, методы, математические модели. – 2007. – № 25. – С. 70-98.
4. Коченков А.И., Толстова Ю.Н. Идеи латентно-структурного анализа Лазерсфельда в современной социологии // Социология: методология, методы, математическое моделирование. – 2003. – № 16. – С. 125-149.
5. Мартышенко Н.С., Мартышенко С.Н. Технологии повышения качества данных в анкетном опросе // Практический маркетинг. – 2008. – №1. С. 8–13.
6. Мартышенко Н.С., Мартышенко С.Н., Кустов Д.А. Средства разработки типологий по данным анкетных опросов в среде EXCEL // Академический журнал западной Сибири. – 2007. – №1. – С. 75–77.
7. Маслак А.А. Измерение латентных переменных в социальных системах. – Славянск-на-Кубани: Издательский центр СГПИ, 2012. – 432 с.
8. Омельченко Г.Г. Теоретико-графовое моделирование сегментации рынка потребителей услуг // Экономика. Право. Печать. Вестник КСЭИ. – 2014. – № 1 (61). – С. 145-154.
9. Поль Р. Гембл, Энтони Марселла, Алан Тапп Глава из книги «Маркетинговая революция» [Электронный ресурс]. – <http://www.marketing.spb.ru/lib-research/segment/revolution.htm?printversion> (дата обращения: 2.12.2014).
10. Понизовкин Д.М. Построение оптимального графа связей в системах коллаборативной фильтрации // Программные системы: теория и приложения. – 2011. – Т. 2. – № 4. – С. 107-114.
11. Татарова Г.Г. Основы типологического анализа в социологических исследованиях: Учебное пособие / Г. Г. Татарова. – М: Издательский дом «Новый учебник» 2004. – 206 с.
12. Ядов В.А. Стратегии и методы качественного анализа данных // Социология: методология, методы, математическое моделирование. – 1991. – № 1. – С. 014-031.

Рецензенты:

Латкин А.П., д.э.н., профессор, директор института подготовки кадров высшей квалификации ВГУЭС г. Владивосток;

Ембулаев В.Н., д.э.н., профессор кафедры Математики и моделирования ВГУЭС, г. Владивосток.