

А.А. ДАВЫДОВ, А.Н. ЧУРАКОВ

СИСТЕМНЫЙ ПОДХОД К АНАЛИЗУ ДАННЫХ МОНИТОРИНГА ОБЩЕСТВЕННОГО МНЕНИЯ

ДАВЫДОВ Андрей Александрович — доктор философских наук, главный научный сотрудник Института социологии РАН, руководитель группы "Анализ социальных систем". ЧУРАКОВ Александр Николаевич - кандидат социологических наук, заведующий кафедрой информационных технологий в социологии Московского государственного социального университета (МГСУ).

Анализ данных мониторинга общественного мнения, т.е. изменения с течением времени долей ответов респондентов по одной или нескольким градациям вопросов анкеты, наиболее часто изучают с помощью так называемого "глазомерного" метода, когда социолог рассматривает данные мониторинга и исходя из здравого смысла и знания изменений социально-экономической и политической ситуации в стране или регионе дает правдоподобное объяснение изменению долей ответов респондентов. Очевидно, что в этом случае истинность интерпретаций трудно проверить, а количество полезной информации, извлеченной из данных при помощи "глазомерного" метода, существенно ограничено.

Анализ данных мониторинга общественного мнения также традиционно осуществляют с помощью статистического подхода, в частности анализа временных рядов. При этом доли ответов (в процентах) респондентов по одной или нескольким градациям одного или различных вопросов анализируют с помощью коэффициентов автокорреляции и кросскорреляции, методов авторегрессии, скользящего среднего, спектрального анализа и некоторых других методов [1]. Напомним, что методы статистического анализа временных рядов направлены на выделение тренда (долговременной составляющей временного ряда), сезонной и циклической компонент, случайной составляющей и ряда других признаков, выявление связи и лагов (запаздываний) между временными рядами и т.д. При данном статистическом подходе истинность полученных результатов основывается на точности описания и предсказания и на возможности содержательно объяснить полученный результат, как правило, с точки зрения здравого смысла, поскольку статистический анализ временных рядов не основан на какой-либо социологической теории, а представляет собой раздел математической статистики, который безразличен к содержательной специфике временного ряда.

С проникновением в эмпирическую социологию компьютерных информационных технологий [2] для анализа динамики используют компьютерные системы извлечения знаний - "data mining", которые методом полного перебора возможных вариантов выявляют различные закономерности в имеющихся эмпирических данных и где истинность полученных результатов определяется только исходя из точности описания и прогноза. Среди систем "data mining" наибольшее распространение получили так называемые "нейронные" сети [3]. Напомним, что "нейронная сеть" состоит из элементов "нейронов", которые размещены по m слоям, обычно трем. Первый слой - входной, второй - скрытый, третий - выходной. "Нейроны" первого слоя получают входные сигналы, преобразуют их по различным правилам, например, с помощью булевой логики, нечеткой логики, весовых коэффициентов, различных формул и т.д. и передают сигнал на другой слой. В процессе обучения на контрольных примерах "нейронной сети" подаются на вход объекты с заданными характеристиками, а на выходе задают принадлежность объекта к определенному классу или какому-либо количественному значению. "Нейронная сеть" изменяет количество "нейронов", их

состояния, количество слоев, правила преобразования для каждого "нейрона" до тех пор, пока точность предсказания принадлежности объекта к заданному классу или какому-либо количественному значению не будет максимальной. При этом неизвестно, по какому правилу сеть точно предсказывает результаты, т.к. "нейронная" сеть - это множество взаимосвязанных различных правил и формул.

Поскольку анализ временных рядов и "нейронные" сети не основаны на какой-либо социологической теории, то при анализе данных мониторинга общественного мнения с их помощью трудно получить теоретически обоснованные выводы, выдвигать новые плодотворные содержательные гипотезы и ставить новые перспективные исследовательские задачи. В этой связи предлагается использовать системный подход к анализу данных мониторинга общественного мнения, поскольку "глазомерный", а также статистический подход и "нейронные" сети проигрывают системному подходу в богатстве возможных теоретических идей и гипотез, количестве содержательных и полезных для социологической теории и практики характеристик динамики систем, широте методов анализа данных мониторинга общественного мнения. Поскольку многие социологи не знакомы с количественным системным подходом, кратко дадим несколько пояснений, необходимых для правильного понимания представленных в данной статье эмпирических результатов [см. также: 4].

Данный подход основан на принципе системности, согласно которому существует зависимость каждого элемента, части, свойства и отношения от его места и функций внутри системы; каждая ее часть является системой, а исследуемая система является частью более общей системы; поведение системы есть результат взаимодействия всех свойств, отношений, элементов и уровней системы, влияния других социальных систем и окружающей среды, а также прошлого состояния системы и ожидаемого будущего. Наряду с принципом системности в системном подходе используются и другие принципы, в частности, принцип соразмерности частей [5], согласно которому в системе существует устойчивое количественное соответствие между размерами частей, необходимое для существования и функционирования системы. Это соответствие может быть выражено в виде прямых пропорциональных закономерностей $y = kx$, где y , x - размеры частей, k - коэффициент пропорциональности, аддитивных закономерностей вида $y = X_1 + x_2 + \dots + x_n$, где y , X_1 , ..., x_n - размеры частей системы, а также других классов закономерностей.

Рассматриваемый подход базируется на принципе детерминизма, т.е. наличии в системе детерминистских законов и закономерностей, т.е. необходимых устойчивых связей между ее элементами и частями. Среди методов выявления закономерностей в системах часто используется метод полного перебора вариантов в уже собранных эмпирических данных. Этот метод обусловлен принципом целостности, согласно которому для успешного анализа и прогнозирования функционирования системы необходимо рассматривать всю систему как единое целое, при учете всех закономерностей, которые обуславливают ее строение и динамику. Кроме того, в силу значительной уникальности каждой системы трудно предварительно сказать, какие количественные закономерности могут наблюдаться в конкретной социальной системе в данный момент времени, а какие закономерности наблюдаться не должны. Истинность полученных результатов определяется здесь прежде всего исходя из соответствия полученных результатов известным системным принципам и общим законам, количественных критериев точности описания и прогнозирования, устойчивости закономерностей в рамках длительного периода. При этом также используется принцип ограниченности, согласно которому в силу принципиальной неполноты наших теоретических и эмпирических знаний о сложной развивающейся системе, ограниченности периода наблюдения и фрагментарности эмпирических данных, имеющих погрешности измерения, теоретических, психологических, культурных, этических и иных ограничениях исследователя, допускается отсутствие содержательных объяснений полученных результатов, а объяснения с точки зрения здравого смысла не считаются обоснованными.

Являются ли данные мониторинга общественного мнения системой? В качестве ее

частей рассматриваются доли ответов респондентов на все вопросы анкеты. Из системного принципа соразмерности частей вытекает, что если в системе существуют устойчивые количественные отношения между размерами частей в системе, то их наличие в эмпирических данных будет эмпирическим свидетельством того факта, что данные мониторинга общественного мнения являются системой.

Мы проанализировали данные мониторинга общественного мнения, проводимого Всероссийским центром изучения общественного мнения (ВЦИОМ) за период 1994—2001 гг. (данные, а также ценные рекомендации по их анализу любезно предоставлены А.И. Гражданкиным - зам. директора ВЦИОМ) с помощью компьютерной экспертно-диагностической системы МАКС (версия 3.1) [6], в которой реализован поиск различных вариантов зависимостей методом полного перебора.

Напомним, что данные мониторинга ВЦИОМ представляют собой распределение ответов респондентов на 32 вопроса, при этом опросы проводятся через каждые два месяца на выборке, репрезентирующей по основным социально-демографическим характеристикам население России. В силу ограниченности данной статьи мы приведем только три класса (1-3) найденных нами устойчивых закономерностей в данных мониторинга общественного мнения ВЦИОМ.

Прямые пропорциональные закономерности. (I) $y = kx$, где y , x - доли ответов респондентов по разным градациям различных вопросов, k - коэффициент пропорциональности.

Аддитивные закономерности. (II) $y = x_1 + x_2 + \dots + x_n$, где y , x_1, \dots, x_n - доли ответов респондентов по разным градациям различных вопросов.

Мультипликативные закономерности. (III) $y = x_1 \times x_2 \times \dots \times x_n$ где y , x_1, \dots, x_n - доли ответов респондентов по разным градациям различных вопросов.

Скажем сразу, что в данных мониторинга общественного мнения ВЦИОМ обнаружены устойчивые для периода 1994-2001 гг. известные системные закономерности (I—III) между размерами частей в системе.

Прямо пропорциональные закономерности (I). Для порогового значения вариации коэффициента пропорциональности k , не превышающего 10%, было найдено 28 прямо пропорциональных закономерностей, часть из которых указана ниже. Отметим, что это составляет 0,2% от всех логических возможных комбинаций двух долей ответов в данных мониторинга. Для удобства описания примера придадим каждой доле ответов на тот или иной вопрос последовательную числовую нумерацию: Как бы вы оценили экономическое положение России? (плохое) - 1; Как бы вы оценили в настоящее время материальное положение вашей семьи? (плохое) - 2; Как бы вы оценили в целом политическую обстановку в России? (напряженная) - 3; Как бы вы оценили в настоящее время материальное положение вашей семьи? (среднее) - 4; Как вы считаете, какое из приведенных ниже высказываний более соответствует сложившейся ситуации? (жить трудно, но можно терпеть) - 5; Что бы вы могли сказать о своем настроении в последние дни? (испытываю напряжение, раздражение) - 6; Как вы думаете, что ожидает Россию в ближайшие месяцы в политической жизни? (некоторое улучшение ситуации) - 7; С каким из следующих суждений о прожиточном минимуме вы бы скорее согласились? (это такой уровень доходов, который обеспечивает человеку скромное, но более-менее приличное существование) - 8; Как бы вы оценили экономическое положение в вашем городе, сельском районе? (плохое) - 9; Если такого рода митинги, демонстрации протеста состоятся, вы лично примете в них участие или нет? (скорее всего, нет) - 10; Как вы думаете, что ожидает Россию в ближайшие месяцы в области экономики? (некоторое улучшение ситуации) - И; Как вы считаете, экономические реформы сейчас нужно продолжать или их следует прекратить? (следует прекратить) - 12. Получены следующие закономерности: (1) = 0,679 x (8); (2) = 0,839 x (9); (3) = 0,991 x (10); (4) = 0,921 x (5); (1) = 0,857 x (3); (5) = 0,776 x (10); (6) = 1,073 x (2); (7) = 1,017 x (11); (6) = 1,498 x (12).

Отметим, что остатки для всех найденных прямо пропорциональных закономерностей распределены нормально, поскольку эмпирические значения статистики Кол-

Количество найденных закономерностей (II) для двух слагаемых при различных величинах абсолютной погрешности

Абсолютная погрешность, %	Весь мониторинг	Мониторинг без градаций "затрудняюсь ответить" и рейтинга политиков	Обобщенные частотные распределения
0,0	0	0	879
0,5	2	0	32653
1,0	26	10	64611
1,5	143	40	96347
2,0	381	75	128119
2,5	908	132	159803
3,0	1762	238	191169

могорова-Смирнова не превышают табличных значений. Это означает следующее. Распределение погрешностей формируется под воздействием очень большого числа взаимно независимых факторов, причем воздействие каждого отдельного фактора мало, равновероятно по знаку и не может доминировать среди остальных. Последовательный характер воздействия случайных факторов на величину погрешности - аддитивный, т.е. влияние каждого случайного фактора суммируется. Иными словами, погрешности приближения можно признать зависящими от многих случайных факторов, а выявленные нами закономерности не являются случайными.

Аддитивные закономерности (II). Полученные нами результаты представлены в табл. 1, где указано количество найденных закономерностей (II) для двух слагаемых при различных величинах абсолютного порога погрешности приближения. Под обобщенными частотными распределениями понимались одномерные частотные распределения ответов респондентов, в котором каждая градация представляла собой сумму абсолютных частот ответов респондентов за весь анализируемый период. Иными словами, данные обобщенные частотные распределения отражают распределение ответов в целом за период 1994-2001 гг. Из табл. вытекает, что на обобщенных частотных распределениях наблюдается больше закономерностей (II) для двух слагаемых. Это может быть связано с тем, что при увеличении объема наблюдений величина абсолютной погрешности уменьшается и существует больше закономерностей, чем при анализе динамики. Здесь просматривается аналогия с законом больших чисел в математической статистике. Напомним, что в соответствии с законом больших чисел сумма большого числа независимых случайных величин утрачивает характер случайной величины и становится близкой к некоторой константе.

Статистический анализ абсолютных погрешностей найденных закономерностей (II) для двух слагаемых за период 1994—2001 гг. показал, что распределение погрешностей в достаточно большом числе случаев приближенно описывается нормальным распределением, поскольку эмпирические значения статистики Колмогорова-Смирнова не превышают табличных значений. Иными словами, погрешности приближения можно признать зависящими от многих случайных факторов, а выявленные нами закономерности не являются случайными.

Мы проанализировали согласованность изменений долей ответов из найденных нами закономерностей (II) для двух слагаемых с помощью коэффициентов кросскорреляции, используемых для анализа временных рядов [1]. Оказалось, что доли ответов в найденных нами закономерностях действительно изменяются согласованно, т.е. наибольшие значения коэффициентов кросскорреляции наблюдались при лаге (запаздывании) равном нулю. Таким образом, имеются основания полагать, что найденные нами закономерности (II) для двух слагаемых описывают не отношения между постоянными величинами или величинами, которые случайным образом изменяются в некотором интервале, а функциональные зависимости.

Количество найденных закономерностей вида (II) при различных величинах абсолютной погрешности и количестве слагаемых

Абсолютная погрешность, %	Три слагаемых	Четыре слагаемых	Пять слагаемых*
0,0	0	0	0
0,5	0	11	98
1,0	10	118	750
1,5	49	404	2585
2,0	146	1028	6406
2,5	376	2321	13360
3,0	803	4868	25531

* Поиск аддитивных закономерностей для пяти слагаемых с целью упрощения вычислений осуществлялся на массиве данных с исключенными пропущенными значениями.

В качестве примера приведены некоторые найденные нами закономерности вида (II) для двух слагаемых при абсолютной погрешности не более 1%. Для удобства пронумеруем. Как бы вы оценили в настоящее время материальное положение вашей семьи? (очень хорошее) - 1; Как бы вы оценили в целом политическую обстановку в России? (благополучная) - 2; Как вы думаете, что ожидает Россию в ближайшие месяцы в политической жизни? (значительное улучшение ситуации) - 3; Как бы вы оценили экономическое положение в Вашем городе, сельском районе? (очень хорошее) - 4; Как бы вы оценили экономическое положение России? (очень хорошее) - 5; Как, на ваш взгляд, будут меняться цены на основные потребительские товары и услуги в ближайшие 1-2 месяца? (будут снижаться;) - 6; Как вы думаете, что ожидает Россию в ближайшие месяцы в области экономики? (значительное улучшение ситуации) - 7. Получено: (1) = (4) + (5); (2) = (5) + (6); (3) = (4) + (7); (3) = (5) + (7); (6) = (1) + (5).

Затем выявили закономерности (II) для трех, четырех и пяти слагаемых без градаций "затрудняюсь ответить" и рейтинга политиков. В табл. 2 представлены полученные результаты. В качестве примера ниже приведены некоторые найденные нами закономерности для четырех слагаемых при абсолютной погрешности не более 1%. Пронумеруем. Что бы вы могли сказать о своем настроении в последние дни? (нормальное, ровное состояние) — 1; В какой мере, на ваш взгляд, заслуживает доверия парламент Российской Федерации? (вполне заслуживает) - 2; Как вы думаете, что ожидает Россию в ближайшие месяцы? (некоторое улучшение ситуации) - 3; Как бы вы оценили экономическое положение России? (очень хорошее) - 4; Как бы вы оценили экономическое положение в вашем городе, сельском районе? (очень хорошее) - 5; Как бы вы оценили в настоящее время материальное положение вашей семьи? (хорошее) - 6; Если говорить в целом, в какой мере вас устраивает сейчас жизнь, которую вы ведете? (по большей части устраивает) - 7; Как бы вы оценили экономическое положение России? (хорошее) - 8; Какую оценку от 1 до 10 вы бы дали председателю кабинета министров? (6) - 9; Какую оценку от 1 до 10 вы бы дали председателю кабинета министров? (9) - 10; Как бы вы оценили в целом политическую обстановку в России? (спокойная) — 11; В какой мере, на ваш взгляд, заслуживают доверия политические партии? (не вполне заслуживают) - 12; В какой мере, на ваш взгляд, заслуживают доверия политические партии (вполне заслуживают) - 13; Как вы думаете, насколько возможны сейчас в городе/сельском районе массовые выступления против роста цен и падения уровня жизни? (маловероятны) — 14; Какую оценку от 1 до 10 вы бы дали президенту России? (10, самая высокая) - 15; В какой мере, на ваш взгляд, заслуживают доверия политические партии? (совсем не заслуживают) - 16; Как

вы считаете, какое из приведенных ниже высказываний более соответствует сложившейся ситуации (все не так плохо и можно жить) - 17. Получено: (1) = (4) + (7) + (9) + (12); (2) = (5) + (8) + (10) + (13); (3) = (6) + (4) + (11) + (13); (14) = (4) + (15) + (16) + (17).

Мультипликативные закономерности (III). Количество найденных закономерностей (III) для различного числа множителей оказалось следующим: 2 (число множителей) - 54 (число закономерностей); соответственно 3 и 231; 4 и 1238; 5 и 5447. Поиск закономерностей проводился на данных мониторинга ВЦИОМ без градаций "затрудняюсь ответить" и рейтинга политиков при абсолютной погрешности в 1%.

В качестве примера ниже приведены некоторые найденные нами мультипликативные закономерности для двух множителей при абсолютной погрешности не более 1%. Что бы вы могли сказать о своем настроении в последние дни? (прекрасное настроение) - 1; Как бы вы оценили в настоящее время материальное положение вашей семьи? (очень хорошее) - 2; Как вы думаете, что ожидает Россию в ближайшие месяцы в области экономики? (значительное улучшение ситуации) - 3; Как бы вы оценили в целом политическую обстановку в России? (благополучная) - 4; Как бы вы оценили экономическое положение России? (хорошее) - 5; В какой мере, на ваш взгляд, заслуживает доверия правительство России? (вполне заслуживает) - 6; Как, на ваш взгляд, будут меняться цены на основные потребительские товары в ближайшие 1-2 месяца? (будут снижаться;) - 7; Как, на ваш взгляд изменились в прошлом месяце денежные доходы вашей семьи по сравнению с ценами? (денежные доходы значительно опережали рост цен) - 8. Получено: (1) = (2) x (6); (2) = (5) x (7); (3) = (2) x (8); (4) = (2) x (3).

Поскольку в данных мониторинга общественного мнения ВЦИОМ на периоде 1994-2001 гг. обнаружены устойчивые известные системные закономерности между размерами частей в системе, то имеются основания считать данные мониторинга общественного мнения ВЦИОМ **системой**. Мы предприняли попытку выявить состояние системы, в которой наблюдаются закономерности (I—III) компьютерной экспертно-диагностической системы МАКС (версия 3.1). Поскольку МАКС проводит измерения более чем по 30-ти различным системным числовым характеристикам по каждому вопросу анкеты, за каждый момент времени и за период времени в целом, выдавая при этом десятки таблиц, графиков, много текстов диагностики на русском языке, то здесь приведены только некоторые полученные результаты. Кроме того, для анализа динамики мы также использовали пакет SPSS, с помощью которого вычисляли коэффициенты автокорреляции.

Как правило, каждая доля ответов наиболее сильно коррелирует с одной предшествующей долей ответов в динамике. Коэффициент автокорреляции имеет лаг (запаздывание) 1, а наибольшая величина коэффициента автокорреляции изменяется от 0,203 до 0,871 по разным градациям ответов. Полученный результат означает следующее. Динамика системы приближенно обладает так называемым Марковским свойством, согласно которому каждое последующее наблюдение зависит только от предшествующего наблюдения и не зависит от более ранних наблюдений.

Размеры частей в данных мониторинга находятся в интервале 0,1 - 81,2%, среднее отношение между всеми долями ответов за взятый период - 2,25, среднее отношение целое/большая часть в распределениях ответов - 2,47, средняя энтропия распределений ответов - 1,86, средняя избыточность распределений ответов - 0,19, среднее отношение для упорядоченных по убыванию долей ответов для всего мониторинга составила 1,618.

Среднее соответствие закону Ауэрбаха за данный период составило 66,16%, что соответствует значительной степени проявления уравнивающего "механизма" в динамике системы. Динамика системы относится к маятниковому типу, что характерно для систем, находящихся в стационарном состоянии, т.е. в установившемся режиме функционирования.

Проанализировав эти и другие результаты, МАКС выдал следующую диагностику

состояния системы. Анализируемая система находится в стационарном состоянии динамического равновесия. Таким образом, имеются основания утверждать, что наблюдаемые закономерности (I—III) присутствуют в системах, которые находятся в стационарном состоянии динамического равновесия. Данное стационарное состояние содержательно выражается в том, что модальные (наиболее часто встречающиеся) ответы респондентов по каждому вопросу свидетельствуют о субъективном неблагополучии граждан России. Так например, за указанный период наибольшая доля ответов респондентов наблюдалась по следующим градациям ответов: испытываю напряжение, раздражение; экономическое положение в городе плохое; экономическое положение России очень плохое; политическая обстановка в России напряженная; в течение ближайшего года никакого улучшения не произойдет и т.д. Иными словами, данные мониторинга общественного мнения ВЦИОМ за данный период отражают мнения респондентов, которые находятся в стабильно неблагоприятной социально-экономической ситуации. Такое стационарное состояние на уровне респондентов обеспечивается соответствующими комбинациями ответов респондентов на разные вопросы анкеты.

Выявленные закономерности могут быть использованы для прогнозирования долей ответов респондентов в данных мониторинга общественного мнения ВЦИОМ, восстановления пропущенных значений и т.д., а также для построения компьютерной модели, в которой доли ответов респондентов по различным вопросам связаны между собой с помощью закономерностей (I—III). Тогда изменяя величину какой-либо одной доли или нескольких долей одновременно, можно проводить компьютерные эксперименты и изучать изменения в данных мониторинга общественного мнения.

Можно с уверенностью предположить, что в данных мониторинга общественного мнения ВЦИОМ существуют и другие, более сложные закономерности, например, следующие: $y = (x_1 + x_2) - x_3$, $y = (x_1 - x_2) + x_3$, $y = (x_1 + x_2) - (x_3 + x_4)$, различные арифметические равенства, например, $x_1 + x_2 = x_3 + x_4$, $x_1 - x_2 = x_3 - x_4$ и т.д., где y , x_1 , x_2 , x_3 , x_4 - доли ответов респондентов на градации различных вопросов. С алгебраической точки зрения, доли ответов, для которых соблюдаются закономерности (II), являются так называемой полугруппой [7] - множеством чисел с операцией сложения, удовлетворяющей закону ассоциативности $(a + b) + c = a + (b + c)$, замкнутые относительно операции сложения, т.е. содержащие вместе с любыми двумя своими элементами их сумму. Алгебраическая теория полугрупп позволяет использовать уже известные теоремы и свойства полугрупп для математического изучения общественного мнения.

СПИСОК ЛИТЕРАТУРЫ

1. Тюрин Ю.Н., Макаров А.Л. Статистический анализ данных на компьютере. М.: ИНФРА - М., 1998.
2. Чураков А.Н. Информационное общество и эмпирическая социология // Социол. исслед., 1998, № 1. С. 35-44.
3. Горбань А.Н., Россиев Д.А. Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996; Ежов А.Л., Шуйский С.А. Нейрокомпьютинг и его приложения в экономике и бизнесе. М.: МИФИ, 1998.
4. Клир Д. Наука о системах: новое измерение науки // Системные исследования. Ежегодник. М.: Наука, 1983. С. 61-85.
5. Сороко Э.М. Структурная гармония систем. Минск: Наука и техника, 1984.
6. Давыдов А.А., Чураков А.Н. Модульный анализ и моделирование социума. М.: ИСАН, 2000.
7. Клиффорд А.Х., Престон Г.Б. Алгебраическая теория полугрупп. Т. 1-2. М.: Мир, 1972