# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 007 ББК 32.812

В.Т. Преслер

## ЭНТРОПИЯ И ИНФОРМАЦИОННЫЙ АНАЛИЗ ПРЕДПРИЯТИЙ. ПРИНЦИПЫ ИНФОРМАЦИОННОГО АНАЛИЗА

В последние годы явно усиливается тенденция использования энтропийных оценок в анализе состояния производственных комплексов и систем [1]. Но при этом упор делается на отдельные, хотя и однотипные, элементы целостной социально-производственной системы. Эти элементы вычленяются из среды функционирования системы и рассматриваются как отдельная самостоятельная система. Такой подход значительно облегчает анализ данных, характеризующих производственное состояние и деятельность элементов (шахты, разрезы, их производственные объединения, угледобывающие районы), однако искажает семантику (содержание) информации в целом и ее количественные характеристики в частности.

Используемые в анализе экономические, технические и технологические показатели (количественная информация – данные о функционирующем объекте анализа, его количественные характеристики), отражают отдельные стороны жизнедеятельности элементов, но не дают представления об их внутренней организованности, потенции, взаимосвязи друг с другом и природной средой, направленности на выполнение целевых установок. Семантика информации играет не менее важную, а может и более важную роль в анализе, чем отдельные количественные характеристики элементов. При этом семантика информации отражает внутреннюю организацию элементов, их взаимосвязь и взаимодействие с внешней средой, системную организацию, управляемость и упорядоченность отдельных элементов и системы в целом, ее устойчивость и способность адаптироваться к изменяющимся условиям среды, вырабатывать целевые установки и эффективно претворять их в жизнь.

Информация в ее количественном и семантическом аспекте объективно отражает функционирование социально-производственного организма. В этом плане анализ производственных систем и отдельных ее элементов с позиций информационных процессов и количественных мер информации обоснован и необходим.

Не следует переоценивать возможности информационного анализа, хотя бы из-за слабости и ограниченности теоретических проработок в этом направлении. Его нужно рассматривать как необходимое дополнение к экономическому и социальному анализу (экономико-социальное моделирование) производственной деятельности, например, угольной отрасли в целом и ее отдельных предприятий в частности. Сфера применения информационного анализа с позиций энтропии в целом ограничивается задачами:

- выявления степени устойчивости системы и ее звеньев и эффективности ее структуризации и степени допустимой свободы ее элементов;
- оценки динамики рыночной среды как множества неравновероятных событий и ситуаций и оценки соответствующего множества стратегий поведения, в той или иной мере удовлетворяющих конкретному проявлению динамики рынка.

В информационном анализе важная роль принадлежит энтропийной мере информации. В нашем случае энтропия, согласно теории информации, отражает меру неупорядоченности системы и ее элементов и неопределенность стратегий ее (их) поведения в рамках среды функционирования. В рамках единого социально-производственного организма, действующего для достижения определенных целевых установок, энтропия, как мера неопределенности, снижается и гораздо меньше, чем у множества однотипных, но разрозненных, отдельно функционирующих элементов. Чем более упорядочено множество, чем более оно системно, тем эффективнее оно функционирует в целом (в смысле достижения целевых установок) и тем меньше ее энтропия.

Энтропия системы в целом определяется тремя составляющими (структурная (системная), элементная и привнесенная энтропия), которые находятся в тесной взаимосвязи, не подчиняющейся закону линейного суммирования: энтропия системы не аддитивна, т.е. не есть сумма трех энтропий. Их связь более сложна. Так, неэффективная структура может многократно усиливать энтропийные всплески отдельных элементов и внешней среды (природная среда, рынок, инфраструктура, социально-политическая среда) и вводить систему в неравновесное состояние. Эффективная структура, обладающая механизмами адаптации, способна гасить энтропийные всплески и проводить в жизнь целевые установки, посредством систематического упорядоченного воздействия на элементы системы в соответствии с процессом адаптации к внешней среде. Здесь надо учитывать, что энтропийное описание процессов в реальной системе всего лишь попытка количественно отразить реальные объективные процессы с позиций информационных мер, что не подменяет реальные социальные и политические механизмы общественного регулирования .

Идеальной системой можно считать такую систему, которая посредством изменения стратегии своего поведения на некотором этапе развития приспосабливается к существующей прогнозируемой динамике внешней среды и стабилизирует свою энтропию на уровне природной. Природная энтропия – объективно неустранима, присуща системе с момента ее организации и эволюционирует по мере пространственно-временного развития системы. Для горнодобывающих предприятий и комплексов это горно-геологический фактор и "груз прошлого" (предыстория воздействия на массив горных пород, проявляемая в последовательном наращивании систем горных выработок и их техногенном воздействии на окружающую среду), который в значительной степени влияет на гидро-, геомеханическое и аэрогазовое состояние шахтного поля и режимы газовой фильтрации в его пределах и смежных полей.

Эти компоненты природной энтропии практически неуправляемы, но их влияние необходимо оценивать и учитывать при организации производственной деятельности. В то же время "груз прошлого" можно постепенно локализовать, включая в горнодобывающую систему шахты модульного типа и выводя из нее старые шахты с протяженной сетью горных выработок. Однако, уменьшая подобным образом природную энтропию, мы не устраняем ее как таковую, только выводим за рамки конкретной системы или ее отдельной подсистемы. Этот фактор остается и продолжает действовать в рамках всего угледобывающего региона.

Идеальная и эффективная системы не тождественны. В отличие от идеальной системы, стабилизирующей энтропию на уровне природной, эффективная система должна обладать гораздо большей энтропией, которая проявляется в достаточной свободе ее элементов и межэлементных связей. Это обуславливает большие возможности по ее перестройке в моменты резких, чрезмерных непрогнозируемых энтропийных всплесков внешней среды. Эффективная система должна содержать механизмы приведения ее в повышенное энтропийное состояние, посредством которых будут гаситься эти энтропийные всплески.

К идеальным системам можно отнести угольную промышленность СССР, но эффективной эта система не являлась. Перестройка, развал страны и последовавшая за этим реструктуризация угольной промышленности разрушили дотоле целостную систему, свели ее к множеству разрозненных, практически не связанных угольных предприятий, предоставленных самим себе, вынужденных действовать в одиночку в непривычных условиях. Налицо энтропийный всплеск, разрушивший структуру системы и поставивший угольные предприятия (элементы системы) перед дилеммой выживания.

Акционирование угольных предприятий и их спорадичный выход на рынок привели к значительному росту информационной напряженности, т.е. росту элементной энтропии. В новых условиях элементная энтропия определяется двумя факторами:

- неопределенность, неизведанность, не изученность рыночных механизмов, обуславливающих потребность в угле и его цену;
- неопределенность в вопросах перестройки самих предприятий, обеспечивающей решение дилеммы выживания и закладывающей базу успешного развития на будущее.

В теории информации и кодирования используется энтропийная мера информации, которая основывается на вероятностной модели событий: получатель информации (сообщения) имеет определенные представления о множестве событий, наступление которых ожидается с некоторыми вероятностями, и общая мера неопределенности (энтропия) характеризуется функцией от совокупности этих вероятностей. Количество информации в сообщении определяется тем, насколько уменьшится эта мера после получения сообщения.

Поясним работу модели на простом примере [2]. Число возможностей выбора из *32*-карточной колоды по одной карте равно *32*. Априори можно предположить, что наши шансы, выбрать некоторую определенную карту, одинаковы для всех карт колоды. Произведя выбор, мы устраняем эту априорную неопределенность. Априорная неопределенность характеризуется числом возможных равновероятных выборов. Если определить количество информации как меру устраненной неопределенности, то полученную в результате выбора информацию также следует оценивать числом возможных выборов *32*. Эта характеристика имеет простой смысл. Чтобы устранить неопределенность, достаточно перебрать все возможные выборы (карты колоды), или зная число выборов, мы однозначно определяем количество информации

Однако в теории информации используется другая количественная оценка энтропии, функционально связанная с предыдущей оценкой. Эта оценка получила название формулы Хартли и показывает, что количество информации, необходимое для снятия неопределенности о системе с равновероятными состояниями, зависит только от количества этих состояний [3]

$$H = log_2 N , \qquad (1)$$

где $N$ – число возможных равновероятных выборов. В качестве единицы измерения используется бит, который определяет два возможных состояния "да/нет" (*0/1*). Отметим, что единица измерения полностью соответствует основанию логарифма, т.е. основной параметр функции (1) и единица измерения согласованы.

В сравнении с числом выборов $N$ оценка (1) имеет более глубокий смысл. В нашем примере $H=5$ характеризует необходимое и достаточное число вопросов, однозначные ответы на которые (да/нет) позволяют со стопроцентной гарантией определить выбранную из колоды карту. Например, для того чтобы установить выбор *дамы пик* следует задать 5 вопросов:

| Вопросы | Ответы |
|---|---|
| Карта красной масти? | 0 – нет |
| Трефы? | 0 – нет |
| Одна из четырех старших? | 1 – да |
| Одна из двух старших? | 0 – нет |
| Дама? | 1 – да |

Стратегия выбора описывается последовательностью из пяти бит *00101*. Подобных стратегий может быть и более одной, но вне зависимости от числа таких стратегий число шагов в них должно равняться *5*. Только в этом случае гарантируется определение выбранной из колоды карты. Использование других (случайных) стратегий поиска (определения) выбранной карты не будут обладать стопроцентной вероятностью, хотя в отдельных случаях могут приводить к успеху за меньшее число шагов. В принципе угадать карту можно и за один шаг, но вероятность при этом будет *1/32*.

В нашем примере число выборов равно степени 2, поэтому число вопросов является целым числом. В случае, когда число выборов не равно степени 2, число вопросов получается нецелым. Так для колоды из *36* карт число вопросов находится между *5* и *6*. В действительности же, чтобы установить выбранную карту, необходимо в ряде случаев задать пять вопросов, а в ряде других случаев шесть вопросов. Множество стратегий поиска состоит из *5* и *6* шаговых стратегий.

Таким образом, оценка (1) характеризует множество неслучайных многошаговых стратегий поиска (действий субъекта в условиях неопределенности), обеспечивающих достижение цели. Оценка однозначно говорит о том, что для устранения неопределенности ситуации необходима *неслучайная многошаговая стратегия поиска* и дает число этих шагов. Однако данная оценка не определяет содержание этих стратегий и специфику конкретных шагов. Применительно к нашему случаю многошаговость следует трактовать как последовательный эволюционирующий выход субъектов на рынок и постепенное расширение своего места в нем, учитывая, что пошаговое воздействие субъекта на рынок, изменяет его энтропию, либо уменьшает, либо увеличивает неопределенность, а, значит, изменяется число возможных выборов. Изменение числа выборов в ходе реализации конкретной стратегии требует постоянной оценки ситуации на рынке и корректировки стратегии воздействия на него.

К. Шеннону принадлежит обобщение оценки (1) на случай неравновероятности выборов [4]

$$h_i = - log_2 \, p_i \ , \qquad (2)$$

где $i$ – номер выбора, $p_i$ - вероятность и $h_i$ – энтропия $i$-го выбора.

Для определения среднего количества информации по всем неравновероятным выборам используется оценка [4]:

$$H = -\sum_{i=1}^{N} p_i log_2 \, p_i \qquad (3)$$

При равновероятных выборах ($p_i=1/N$ ) оценка (3) приводится к виду (1).

В работе проф. Логова А.Б. и др. [1], посвященной информационному анализу состояния угольного комплекса Кузбасса, вводится иная, отличная от классической шенноновской, оценка количества информации для отдельного элемента комплекса

$$f_i=- d_i \, ln \, (d_i \,), \; i=1 \div N \qquad (4)$$

где $i$ - номер элемента комплекса (производственное объединение), $d_i$ - доля элемента, вклад элемента в некоторую суммарную характеристику комплекса (например, добычу угля, доход, численность).

Рассмотрим три аспекта проявления этих отличий.

**<u>Аспект 1.</u>** Шенноновская трактовка меры энтропии опирается на *случайную событийную модель*, которая характеризуется двумя факторами – числом возможных событий (выборов), и вероятностями свершения этих событий, что и отражается в оценках (2, 3). Авторы [1] предполагают, что показатели элементов (добыча, доход, численность) соответствуют понятию распределения. При этом понятие *распределение* только констатируется, не определяются его характерные свойства и признаки, обосновывающие принципиальную возможность его использования для измерения количества информации.

В рамках жестко определенного множества (в работе [1] это 9 производственных объединений) распределение некоторого показателя (например, добыча) не является случайной событийной моделью и не тождественно функции распределения некоторой случайной величины. Естественно, целенаправленная деятельность объединений в анализируемом периоде в информационном аспекте привела к устранению априорной неопределенности, существовавшей в начале периода. Однако ее устранение определяется не совокупностью результирующих показателей, а ресурсными затратами на добычу тоны угля (технические, технологические, людские, управленческие, интеллектуальные, экологические), ушедших на преодоление горно-геологического фактора и "груза прошлого", характеризующих сопротивление природной среды изъятию сырья.

Распределение показателя можно трактовать как *квази-случайный процесс* только в том случае, если рассматривать множество различных групп (от 2 до 9), содержащих различные комбинации реализации этого показателя по объединениям. В этом случае долю элемента в суммарном весе элементов, образующих отдельную группу, можно рассматривать как реализацию этого квази-случайного процесса, определяющего долевое участие отдельного элемента, его вес в некоторой квази-случайной совокупности элементов. Поскольку доля отдельного элемента в этой квази-случайной совокупности определяется вошедшими в нее элементами (она различна для разных совокупностей), то событием можно считать попадание элемента в совокупность, а долю интерпретировать как вероятность наступления этого события. Устранение априорной неопределенности для данного показателя означает выбор из множества одной группы для информационного анализа вошедших в нее элементов. Применение оценки (3) при таком подходе позволяет определить среднее количество информации, соответствующее выбору структуры для информационного анализа совокупности элементов по их отдельным показателям.

**Аспект 2**. По внешнему виду оценка (4) **похожа** на оценку (3). Отличие в основании логарифмов только количественно изменяет размерность оценок поскольку $log_2 p=ln(p)/ln(2)$. В [3] для оценки (3) с натуральным логарифмом используется единица измерения *нат* в отличие от используемой авторами [1] единицы *нит* для оценки (4). Однако удобней использовать классическую единицу измерения *бит* (*1 бит = 1/ln(2) нит*). Более существенны **смысловые отличия** величин, входящих в обе оценки.

Величины $p_i \, log_2 \, p_i$ характеризуют *энтропийный вклад отдельных событий в среднюю оценку количества информации*, которая, в свою очередь, характеризует в среднем то количество информации, которую мы получим после устранения априорной неопределенности, связанной со свершением отдельного события из их множества. Энтропийный вклад элемента не тождественен его энтропии - он всегда меньше индивидуальной энтропии элемента. В общем случае события могут быть независимыми (представлены своими безусловными вероятностями) и зависимыми (представлены своими условными вероятностями), но в любом случае значения вероятностей лежат в диапазоне [0, 1]. Для зависимых событий, представленных условными вероятностями, необходим пересчет этих вероятностей в их абсолютные вероятности. Пересчет производится в соответствии с событийной взаимосвязью и абсолютными вероятностями независимых событий, определяющих события зависимые.

*Доля* или вес отдельного элемента является *функцией трех факторов*: а) процесса отбора элементов в анализируемое множество (конкретизирует элементы и их число), б) суммарного значения показателя в отобранном множестве, в) значения показателя по отдельному элементу. В отличие от доли, чувствительной к фактору отбора элементов, вероятность события не зависит от того, какие из них отбираются для анализа, хотя неполнота событийного множества влияет на энтропийные оценки анализируемого процесса. Ограничение для элементных долей

$$\sum_{i=1}^{N} d_i = 1 \cdot \qquad (5)$$

даже при независящих друг от друга значениях показателя по элементам, делает доли элементов зависимыми (уменьшает степень свободы элементов). В данном случае, если воспользоваться механизмом случайного определения долей, то 1-ая доля $d_1$ выбирается в интервале [0, 1], 2-ая - в интервале [0, $1-d_1$], 3-я – в интервале [0, $1-(d_1+d_2)$] и так до ($N$-1)-ой доли, последняя определяется из (5).

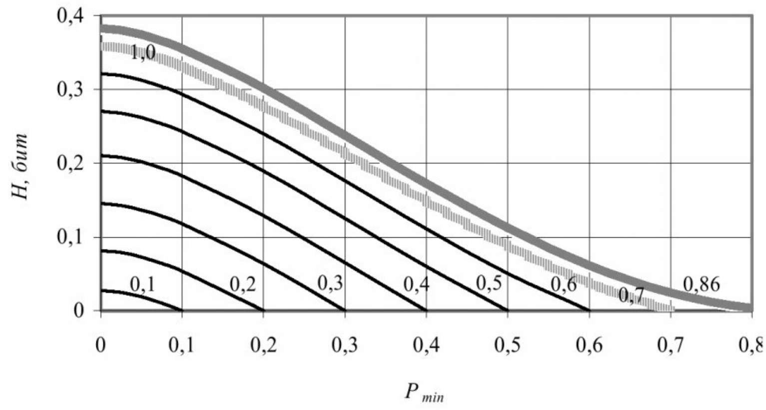


*Рис.1. Энтропия событий, вероятность которых монотонно растет в диапазоне вероятностей [$p_{min}$, $p_{max}$] для различных значений $p_{max}$*

В отличие от дискретного распределения долей в рамках множества элементов, событийная модель допускает непрерывное распределение событий, которое характеризуется монотонным непрерывным изменением их вероятности. В этом случае для оценки количества информации воспользуемся интегральным представлением энтропийной меры.

Рассмотрим бесконечное множество событий, вероятность которых монотонно возрастает в интервале [$p_{min}$ , $p_{max}$]. Тогда, согласно оценке (3), переходя к интегралу, получим среднюю интегральную меру количества информации

$$H = -\frac{1}{p_{max} - p_{min}} \times \qquad (6)$$

$$\times \int_{p_{min}}^{p_{max}} p \cdot log_2 p \, dp =$$

$$= \frac{1}{2ln2}\begin{bmatrix} p_{max}^2 \left(0,5 - log_2 p_{max}\right) - \\ - p_{min}^2 \left(0,5 - log_2 p_{min}\right) \end{bmatrix}.$$

Среднее количество информации, извлекаемое из событий с распределением вероятностей в интервале [0, 1] ($H$=*0,36 бит*), равно сумме средних количеств информации, извлекаемых из двух множеств событий с распределениями вероятностей в интервалах [0, 0,5] ($H$=*0,215 бит*) и [0,5, 1] ($H$=*0,145 бит*). Однако сами эти средние количества информации не равны, причем в полтора раза больше информации извлекается из событий в интервале [0, 0,5]. На рис. 1 приведены графики изменения энтропии в зависимости от минимальной вероятности $p_{min}$ при различных значениях максимальной вероятности $p_{max}$. Как видно, кривая энтропии достигает своего максимального уровня при $p_{max}$ =0.86, а затем снижается до уровня кривой энтропии при $p_{max}$ =0.7. Оценка (6) в интегральной форме обобщает оценку (3) для непрерывного и неравномерного распределения вероятностей. Она позволяет

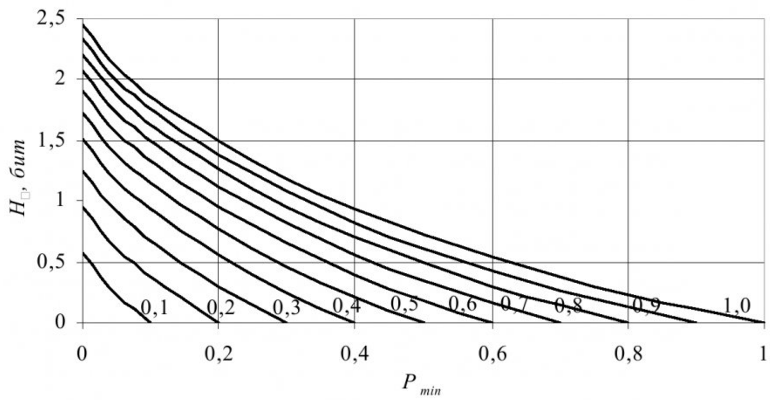


*Рис. 2. Зависимость энтропии последовательности событий H от априорной вероятности событий* $p_{min}$ *для различных значений* $p_{max}$

оценить среднее количество информации, которое будет получено при свершении одного из событий в интервале [$p_{min}$, $p_{max}$].

Рассмотрим некоторый непрерывный процесс, применяемый к множеству равновероятных событий, и состоящий в том, что после свершения некоторых событий они как бы удаляются из множества, а вероятность оставшихся при этом событий возрастает. Оценим количество информации, извлекаемое в ходе процесса, который соответствует наращиваемой сумме событий.

Смысл процесса легко уловить, если воспользоваться его дискретной интерпретацией (колодой из 32 карт). Априорная вероятность выбора карты 1/32. После идентификации выбранная карта удаляется из колоды. Вероятность выбора следующей карты возрастает и составит 1/31. Процесс выбора, идентификации и удаления продолжается до последней карты из колоды, вероятность выбора которой равна 1. Для определения суммарной энтропии последовательного свершения событий воспользуемся суммой индивидуальных оценок информации (2) в интегральной форме

$$H_{\Sigma} = -\int_{p_{min}}^{p_{max}} log_2 p \, dp =$$

$$= \frac{p_{max} - p_{min}}{ln2} + p_{max}\left(1 - log_2 p_{max}\right) - p_{min}\left(1 - log_2 p_{min}\right), \qquad (7)$$

где $p_{min} \geq 0$ - априорная вероятность событий, $1 \geq p_{max} \geq p_{min}$ - конечная вероятность событий, после свершения некоторого их множества. Сред-

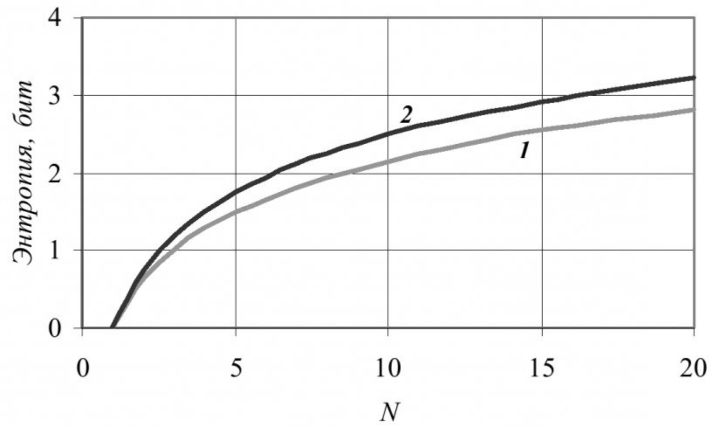


*Рис. 3. Зависимость энтропии, приходящейся на реализацию отдельной доли (1 -* $\overline{F}$ *) и суммы долей (2 -* $\overline{F}_{\Sigma}$*), от числа элементов*

няя мера энтропии для процесса последовательного свершения событий составит

$$\overline{H}_{\Sigma} = H_{\Sigma} \Big/ \left( p_{\max} - p_{\min} \right). \quad (8)$$

Значения средней меры энтропии для различных значений априорной и конечной вероятностей:

$p_{min}=0$, $p_{max}=1$,
$\hat{H}_{\Sigma}=1/\ln 2=1.45$ *бит;*

$p_{min}=0$, $p_{max}=0.5$,
$\hat{H}_{\Sigma}=1+1/\ln 2=2.45$ *бит;*

$p_{min}=0.5$, $p_{max}=1$,
$\hat{H}_{\Sigma}=-1+1/\ln 2=0.45$ *бит.* (9)

Как видно из (9), энтропия процесса последовательного свершения событий значительно (в несколько раз, а в отдельных случаях на порядок) превышает энтропию событий с монотонно возрастающей вероятностью. На рис. 2 приведены графики, характеризующие изменчивость энтропии последовательного свершения событий в зависимости от априорной вероятности событий $p_{min}$ для различных значений $p_{max}$. Как видно, уровень энтропии непрерывно, но замедленно возраста-

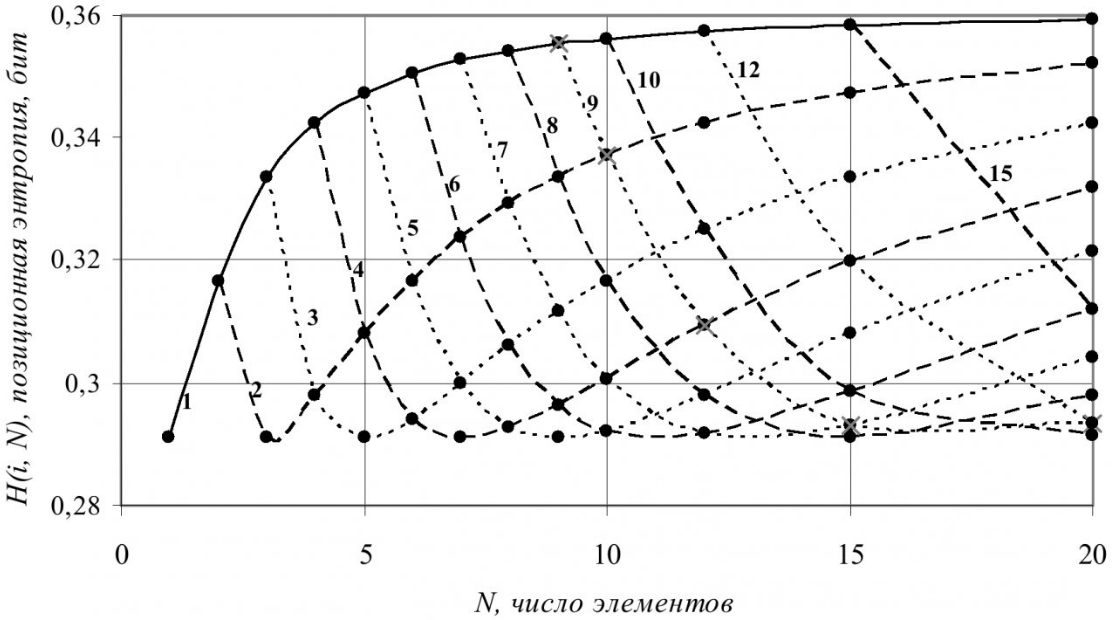


*Рис.4. Распределение позиционной энтропии в зависимости от числа элементов для монотонно возрастающего номера позиции i = 1÷20*

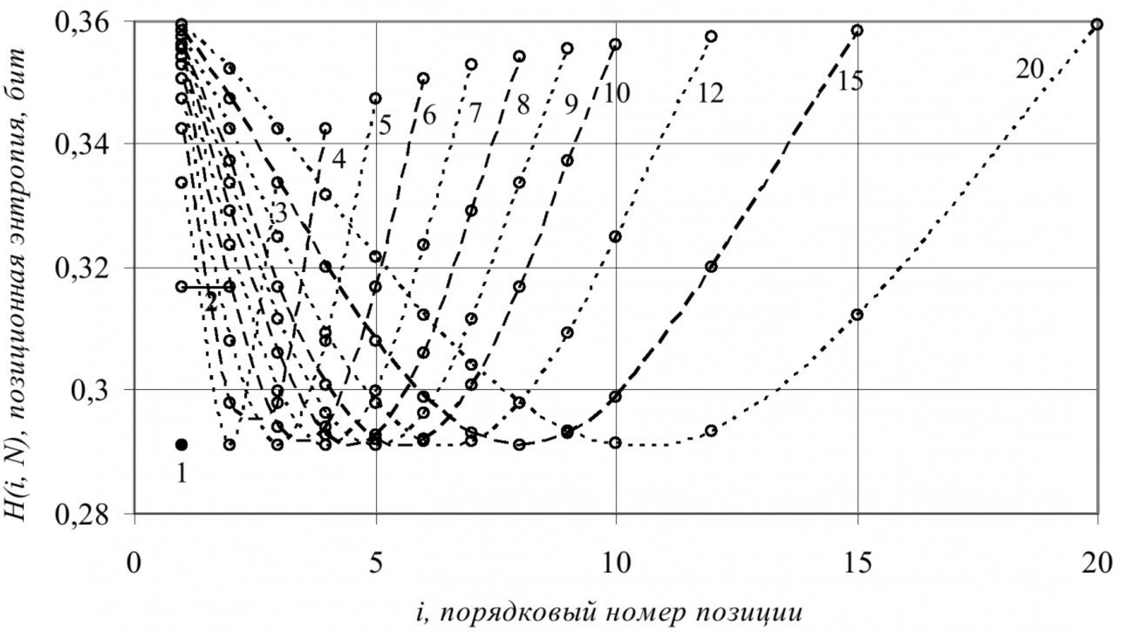


*Рис. 5. Распределение позиционной энтропии в зависимости от номера позиции для монотонно возрастающего числа элементов N = 1÷20*

ет с ростом $p_{max}$.

Возвращаясь к доле элемента, определим среднее количество информации в смысле оценки (4), полагая равномерное увеличение долей элементов $\{ d_i=\Delta \cdot i \ , i=1 \div N\}$. Тогда с учетом (5) получим уравнение связи для шага изменения доли $\Delta$ и числа элементов $N$ и наибольшую долю $d_N$:

$$\Delta = \frac{2}{N(N-1)}, \quad d_N = \frac{2}{N+1}. \quad (10)$$

Среднее количество информации, приходящееся на реализацию некоторой доли элемента, оцениваем по аналогии с (3) и согласно (4):

$$\overline{F} = \sum_{i=1}^{N} f_i = = ln2 \left[ log_2 \frac{N(N+1)}{2} - \frac{2}{N(N+1)} \sum_{i=1}^{N} i \cdot log_2 i \right], бит \quad (11)$$

а среднее количество информации, приходящееся на сумму реализаций долей, оцениваем в соответствии с индивидуальными оценками (2) по формуле:

$$\overline{F}_\Sigma = \frac{1}{N} \sum_{i=1}^{N} ln\ d_i = ln2 \left[ log_2 \frac{N(N+1)}{2} - \frac{1}{N} \sum_{i=1}^{N} log_2 i \right], бит \quad (12)$$

Формулы (11, 12) отражают связь энтропии долей с числом элементов в случае монотонного равномерного роста долей в авторской трактовке [1] (рис. 3). Как видно, энтропия реализации суммы долей превышает энтропию реализации отдельной доли.

Рассмотрим еще один подход, позволяющий перейти к вероятностной трактовке доли элемента. Разделим диапазон ее возможных значений $0<d<1$ на $N$ равных позиций размером $1/N$, где $N$ - число элементов. Под событием будем понимать занятие долей $d$ некоторой позиции, вероятность которой будем оценивать выражением

$$p_i(d)=1-|d-\eta_i| \ , \quad \eta_i =(i-0.5)/N , \quad 0.5/N \le \eta_i \le (N-0.5)/N . \quad (13)$$

Тогда среднее количество информации $H_i=H(i,N)$, определяющее позиционную энтропию событий, состоящих в том, что доля займет $i$-ю позицию, оценивается в интегральной форме по аналогии с (7)

$$H_i = -\int_0^1 p_i(d) log_2 p_i(d) d d = = \frac{1}{2ln2}\left(0,5+\eta_i-\eta_i^2\right) + +0,5\left[(1-\eta_i)^2 log_2(1-\eta_i)+\eta_i^2 log_2 \eta_i\right] \quad (14)$$

На рис. 4 и 5 приведены графики, иллюстрирующие распределение позиционной энтропии в зависимости от числа элементов и номера позиции. Как видно, энтропия колеблется в малом диапазоне, равном 0,07 *бит*, имеет минимумы на уровне 0,29 *бит*, симметрично распределена по позициям относительно средней позиции, в которой достигает минимума (рис.5.). Ее распределение по числу элементов носит несимметричный вогнутый характер с крутым спадом и пологовыпуклым восхождением (рис.4.).

**Аспект 3**. Проанализируем более детально оценки (3) и (4). Как уже отмечалось компоненты $u=p \cdot log_2 p$ (индекс опущен), входящие в оценку (2), характеризуют энтропийный вклад события, имеющего вероятность $p$, в среднюю меру. Энтропийный вклад существенно отличается от индивидуальной меры энтропии (2). Так в точке экстремума $p=1/e \sim 0.37$ вклад максимален и равен 0,53 бит. Сравним, индивидуальная мера энтропии события, имеющего вероятность 0,37, равна 1,44 бит, что почти втрое превышает энтропийный вклад.

На рис.6 приведен график этого вклада (сплошная линия). Как видно из графика функция вклада имеет несимметричный относительно точки экстремума вид, а с уменьшением и ростом вероятности событий относительно этой точки вклад падает и становится равным нулю для событий невозможных ($p=0$) и предопределенных ($p=1$). Наибольшее количество информации извлекается из событий, группирующихся в области экстремума $p=0,37$. Это обстоятельство имеет глубокий физический смысл. События как маловероятные, так и сильно вероятные легко предсказуемы, они почти никогда не происходят, или происходят почти всегда. События со средним уровнем вероятности трудно предсказуемы, их прогноз в наибольшей степени не определен, вероятен.

По аналогии с энтропийным вкладом, определяемым функцией $u$, будем называть долевым вкладом, определяемым функцией $f=-d \cdot ln\ d$, вклад доли элемента в некоторую меру, суть которой еще требуется установить. На рис.6 долевой вклад отмечен пунктирной линией с выделенными точками, смысл которых только в том, чтобы показать дискретный характер доли. Несмотря на внешнюю схожесть функций $u$ и $f$

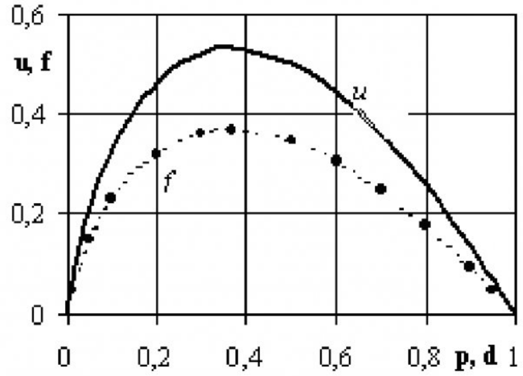


*Рис. 6. Энтропийный и долевой вклады*

(совпадают с точностью до коэффициента $ln2$), они, естественно имеют различный смысл, который определяется не столько дискретным характером доли, сколько ограничением (5), сковывающим свободу ее позиционирования, т.е. способностью занимать различные позиции в диапазоне [0, 1].

Позиционирование доли в рамках различных комбинаций элементов (от 2 до $N$) существенным образом зависит от величины доли. Чем больше величина доли элемента во всем рассматриваемом множестве элементов, тем меньше у нее возможностей позиционирования и, наоборот, чем меньше величина доли, тем больше возможностей позиционирования. Так доля, по величине близкая к 1, будет устойчиво занимать практически одну и ту же позицию в различных комбинациях элементов, а доля, по величине близкая к 0, в зависимости от величин других долей может позиционироваться в пределах диапазона [0, 0.5]. В общем случае доли, занимающие положение между наименьшей, близкой к нулю, и наибольшей, близкой к единице, могут позиционироваться в диапазоне [$d$, 1], если их величины существенно отличаются от наименьшей доли.

С другой стороны, чем ближе доля к 1, тем более ощутимо доминирование соответствующего элемента в информационном анализе множества, а чем ближе доля к 0, тем незначительней. Таким образом, доминирование одной доли приводит к вырождению информационного анализа как анализа совокупности элементов, а малая величина доли определяет ее незначимость и возможность игнорирования в анализе. Обе эти ситуации требуют вывода соответствующего элемента из их множества с целью создания условий для проведения анализа оставшихся элементов.

Из проведенного анализа следует, что функция $f$ может претендовать на роль оценки долевого вклада отдельного элемента в информационный анализ их совокупности. Чем меньше или чем больше этот вклад (по сравнению с оптимальным), тем менее значим элемент, его внесший, в информационном анализе. В данном случае функция $f$ выступает критерием значимости элемента в информационном анализе их совокупности

$$f=-d\ ln(d)\,. \qquad (15)$$

Экстремальный характер критерия значимости (оптимальная величина доли $d=0{,}37$) четко ограничивает рамки анализируемой совокупности и позволяет установить необходимое число ее элементов и распределение их весов, при котором будет обеспечена наибольшая их значимость в анализе.

Рассмотрим случай *неравномерного распределения долей*. Так, если оптимальная доля одного элемента в совокупности составляет 0.37, то оптимальная доля следующего элемента в совокупности оставшихся элементов (элемент с установленной долей выводится из множества) также должна составлять 0.37, что соответствует величине доли элемента в полном их множестве $d$=0.37·0.63=0.23. Для шести элементов, доля которых оптимальна в убывающей совокупности, получим оптимальное распределение их долей

$$0.37 : 0.,23 : 0.15 : 0.09 : 0.06 : 0.04. \qquad (16)$$

Такое распределение не перекрывает весь диапазон [0, 1], но обеспечивает 94 % его перекрытия. Для пяти элементов обеспечивается 90 %-е перекрытие диапазона. Данному распределению долей в полном множестве элементов соответствует распределение значений критерия значимости (15):

$$0.37 : 0.33 : 0.28 : 0.22 : 0.17 : 0.13. \qquad (17)$$

Доля 1-го элемента в распределении примерно в три раза выше доли 6-го элемента. Если колебания значений критерия значимости находятся в диапазоне 10-20 %, то колебание долей в диапазоне 33-38 %, что в 2-3 раза превышает колебание критерия значимости. Отметим, что срединой по данному распределению доле 0,15 соответствует величина критерия значимости 0,28, что всего на 10 % превышает среднее значение по распределению критерия значимости 0,25 и соответствует минимальной величине его колебаний.

Полученное распределение долей и соответственно значений критерия значимости имеет глубокий смысл. Во-первых, есть явно выраженный лидер, но не монополист, поскольку сфера его влияния значительно

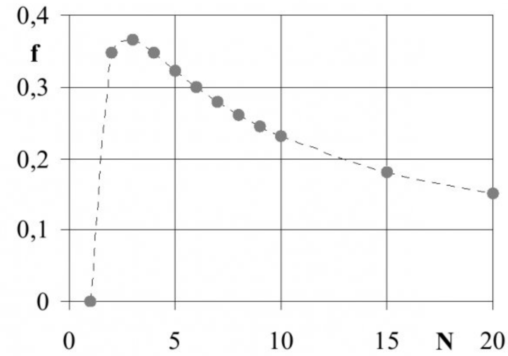


*Рис.7. Критерий элементной емкости*

меньше половины совокупного результата деятельности. Во-вторых, следующая за лидером пара элементов, в силу равенства их суммарной доли доле лидера и в случае их согласованных действий либо объединения, способна составить ему конкуренцию, так же как и любая последующая пара - предыдущему элементу. В третьих, объединение лидера с третьим элементом или с двумя предпоследними (4-м и 5-м) обеспечивает ему подавляющее преимущество (0.52 – совместная доля) по сравнению с объединением 2-го и 3-го элементов (0.38 – совместная доля), объединение с 4-м – значительные преимущества, а объединение с 5-м или 6-м - несущественные преимущества. Такое распределение долей создает необходимые условия ***устойчивого равновесного состояния*** на рынке товаропроизводителей.

Рассмотрим случай *равномерного распределения долей*. В предположении равенства долей справедливо соотношение $d=1/N$, а критерий значимости (15) преобразуется в критерий элементной емкости множества

$$f=ln(N)\,/N\ . \qquad (18)$$

Экстремальный характер критерия (рис. 7) с точностью до целого числа указывает на оптимальное число элементов множества. В соответствии с положением экстремума оно равно *3*. Для трех элементов доля каждого составит 0,33(3), немного меньше оптимального значения 0,37. Для шести элементов (случай неравномерного распределения) значение критерия по сравнению с экстремальным уменьшается менее чем на 20 %. Если положить доли двух элементов равными 0,37, то доля 3-го элемента составит 0,26.

В экономическом рыночном аспекте такое распределение долей в отличие от неравномерного распределения создает условия ***неустойчивого равновесного состояния*** на рынке товаропроизводителей. Это связано с тем, что при реализации элементом стратегии расширения своей сферы влияния, два других элемента могут объединиться, чтобы этого не допустить. С другой стороны, учитывая этот факт, первый элемент может пойти на некоторые уступки в сфере своего влияния, чтобы заручиться частичной поддержкой элемента с меньшей долей, при проведении экспансии в сферу влияния второго элемента. Экспансия элемента с большей долей в сферу влияния элемента с меньшей долей неизбежно приводит к его объединению (совместным действиям) с другим элементом.

Ситуация неустойчивого равновесия отличается энтропийным характером и ее энтропия возрастает с ростом числа элементов, поскольку возрастает неопределенность во взаимодействии (противодействии) элементов. В случае трех элементов множество обладает меньшей энтропией, а значит и большей определенностью во взаимодействии элементов. Множество из двух элементов обладает еще меньшей энтропией (мера равна 1 *бит*). Здесь возможны два состояния - либо договор о разделе сфер влияния, либо конфронтация (исход борьбы не определен).

В физическом аспекте, в условиях противодействия двух сил, наличие третей силы, направление действия которой не совпадает с направлением действия этих сил, способно изменить вектор действия одной или сразу двух сил и тем самым предотвратить последствия их столкновения. Аналогично проявление третьей силы в политической сфере. Ее наличие позволяет сгладить последствия столкновения крайних партийных интересов и придать политическим процессам некоторую устойчивость, хотя и неравновесного свойства. Поскольку у третьей силы есть и свои интересы, то в соответствии с ними будет изменяться ее воздействие на процесс и, как следствие, будет смещаться точка равновесного состояния трех сил.

В математическом аспекте три точки, через которые можно построить кривую, описывающую нелинейный процесс, и определить его динамику гораздо предпочтительней двух и даже четырех точек. Две точки позволяют выявить только линейную тенденцию, а четыре точки, хотя и позволяют извлечь больше информации, но в то же время из-за наличия точек перегиба имеют большую неопределенность в проведении через них различных кривых. С увеличением числа точек растет и энтропийный характер процесса интерполяции.

Таким образом, завершая анализ отличительных свойств шенноновской формулировки количества информации и оценки [1] для отдельного элемента из их множества, можно сделать следующий вывод.

*Введенная авторами [1] оценка (4) характеризует долевой вклад элемента в информационный анализ их совокупности и играет роль критерия значимости (15), согласно которому устанавливается их оптимальное неравномерное распределение, обеспечивающее устойчивое равновесное состояние множества взаимодействующих элементов.*

Проанализируем возможности использования критерия значимости (15) для ранжирования элементов множества по их доле в общей совокупности элементов. В аспекте ранжирования критерий (15) имеет как существенный недостаток, так и существенное преимущество.

Недостаток состоит в ограниченности его использования в диапазоне изменчивости доли. Его можно использовать только для долей, величина которых меньше пороговой величины $d_{lim}=1/e=0,37$. Превышение этого порога приводит к падению значения критерия, а значит и к падению роли элемента в их совокупности, что противоречит росту долевого вклада в совокупность элементов. Данное обстоятельство отмечено авторами работы [1], однако их замечание о возможности преодоления этой ситуации посредством выбора суммарного показателя не спасает положение. В рамках единого интегрального анализа ограниченного множества решить эту задачу невозможно, поскольку его суммарный показатель жестко определен абсолютными значениями показателей по входящим в него элементам и не может подвергаться никаким модификациям.

Единственный выход из такой ситуации – вывод чрезмерно выделяющегося элемента из данного множества и перевод его в некоторое другое множество, в котором он становится соразмерным с его элементами, т.е. его доля чрезмерно не выделяется. После вывода из анализируемого множества выделяющегося элемента, естественно, изменяются доли оставшихся элементов. В этом случае может оказаться, что в рамках претерпевшего изменения множества, вновь появится элемент, доля которого превысит порог, и, значит, вновь появится необходимость в его выводе из множества. Отсюда следует неизбежный вывод.

*Анализ должен быть структурирован по множествам соразмерных элементов, количество которых ограничено (не более 6). Принципы структурированности, соразмерности элементов в анализируемом множестве (узле анализа) и ограниченности числа элементов в узле анализа определяют структуру информационного анализа, а также распределение элементов по его узлам (узловым точкам).*

В этом явно прослеживается диалектика анализа, когда недостаток, свойственный интегральному анализу большой совокупности несоизмеримых элементов, устраняется их дифференциацией по множеству соизмеримых совокупностей элементов, в рамках которых больше возможностей по углубленному и детализированному изучению их свойств и особенностей.

Преимущество критерия (15) перед непосредственным анализом долей элементов состоит в разнесении малых и в сближении больших, но меньше пороговой, долей. Разнесение малых как бы повышает значимость "малого" в анализе, а сближение больших долей как бы выравнивает "большое", делает их менее различимыми в анализе. Естественно, динамику критерия в непрерывном случае полностью характеризует его производная $f'=1+ln(d)$ , темп изменения которой падает по мере роста доли. Однако для дискретного случая предпочтительней анализ изменения разности значений критерия от разности долей элементов.

Рассмотрим дискретную динамику критерия на основе сравнительного анализа двух элементов, доли которых отличаются на величину $\Delta d$. Если доля одного из элементов составляет $d$, то разность критерия (15) (дифференциал $df$ берем по абсолютной величине) для двух элементов составит

$$df = \left| d \cdot lnd - (d + \Delta d) \cdot ln(d + \Delta d) \right| . \qquad (19)$$

Динамику разности (19) иллюстрирует рис. 8, откуда видно, что темп изменения критерия с ростом доли падает, а с ростом разности между долями двух элементов - возрастает. Такой характер поведения критерия обеспечивает рост отклонений в группе элементов с малыми долями и уменьшение отклонений в группе элементов с большими долями. Тем самым обеспечивается выравнивание дисперсии крайних групп. В конечном итоге, выравнивание на уровне критерия дисперсии крайних групп приводит к значительному уменьшению дисперсии всего множества элементов по сравнению с дисперсией их долей.

В работе [1] для ранжирования предприятий по ряду показателей введен среднекомплексный критерий

$$\bar{f}_i = - \sum_{j=1}^{n} d_{ji} ln\, d_{ji} \Big/ n , \qquad (20)$$

где индекс $1 \le j \le n$ характеризует показатели элемента (предприятия), а индекс $i$ распределение показателя по элементам.

Использование оценки (20), как и прежде, ограничивается пороговым значением доли $d_{ji} < d_{lim}$ . Кроме этого, на сами показатели накладывается ряд семантических ограничений.

Во-первых, они должны быть все результирующего типа, т.е. отражать производственно-экономическую и техническую деятельность предприятия в целом, а не просто констатировать отдельные ресурсы или средства.

Во-вторых, они должны быть однонаправленного действия, т.е. их рост должен соответствовать положительной оценке деятельности предприятия, а падение - отрицательной оценке.

В-третьих, показатели должны быть в большой степени независимы, т.е. функционально разделены, иметь как можно меньше «пересекающихся» аргументов.

Однако, использованные в

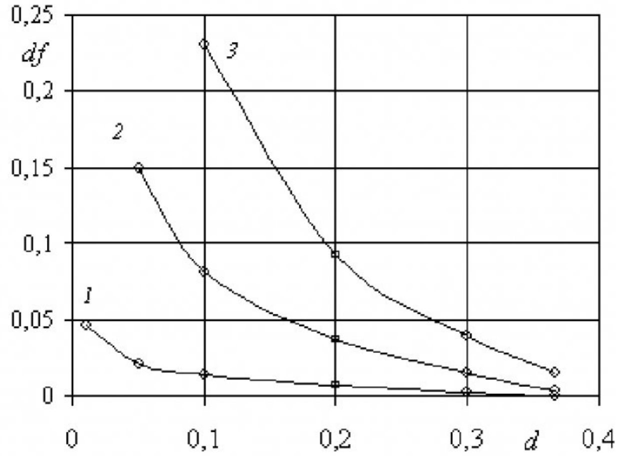


*Рис. 8. Динамика критерия значимости*
*(1 -Δd=0,01; 2 -Δd=0,05; 3- Δd=0,1)*

комплексном критерии (20) показатели (добыча, доход, численность), не удовлетворяют семантическим ограничениям. Результирующие показатели "добыча - доход" функционально связаны через цену угля. При единой цене на уголь доля дохода будет полностью соответствовать доле добычи, т.е. показатель дохода вырождается в показатель добычи. Нерезультирующий показатель "численность" может действовать как однонаправлено с показателями добычи и дохода, так и наоборот, т.е. увеличения добычи и дохода можно добиться и при меньшей численности работников.

Пусть при одной добыче два предприятия отличаются доходом и численностью. Доля дохода 1-го предприятия превышает долю дохода 2-го на некоторую величину $\Delta d_1$, а доля 2-го предприятия по численности превышает долю 1-го на величину $\Delta d_2$. С точки зрения критерия (20) 2-ое предприятие может оказаться предпочтительней 1-го, если $\Delta d_2 > \Delta d_1$. Так как превышение доли по численности больше превышения доли по доходу при равных, например, долях дохода 1-го предприятия и численности 2-го предприятия. Однако с экономической точки зрения 1-ое предприятие явно предпочтительней 2-го, поскольку обеспечивает больший доход при меньшей численности.

Для сопоставительного информационного анализа деятельности предприятий в большей степени подходят относительные комбинированные, а не абсолютные показатели. Абсолютные показатели, например добыча угля, в большей мере определяют место предприятия в сфере рынка товаропроизводителей, т.е. занимаемую нишу, в то время как относительные комбинированные показатели характеризуют производительность труда, эффективность производства единицы продукции и т.д.

Требования оптимальности доли и оптимальности критерия емкости определяют структурированность информационного анализа. Не вызывает сомнения тот факт, что каждое предприятие, в том числе и угольное, в соответствии со своими производственно-экономическими показателями занимает определенную нишу в сфере товаропроизводителей. Определение специфических особенностей и общих закономерностей существования различных ниш товаропроизводителей, анализ их возможностей по улучшению и расширению своей деятельности в пределах занимаемой ниши и возможностей экспансии более высокоуровневой ниши позволяют выделить стратегические направления инвестирования предприятий с целью получения инвесторами определенных выгод. Выгоды могут быть как финансового, так и социального плана. Таким образом, определение стратегий инвестирования предприятий и выявление их потенциальных возможностей является доминантой структуризации информационного анализа. Такая постановка задачи выходит за рамки ранжирования предприятий по показателям их деятельности. Для того чтобы расставить предприятия по рангу в соответствии с их показателями не требуется структуризация анализа, да и смысла в просто ранжировании немного, поскольку значимо различны ниши, занимаемые ими.

Требование *оптимальности доли*, т.е. ее вхождение в область экстремума критерия значимости, может быть обеспечено только для нескольких элементов из их множества (2-3 элемента). Это обстоятельство требует иерархической структуризации самого информационного анализа с выделением пунктов или узлов анализа, которые регламентируют количество элементов, порядок и ступенчатость (многоуровневость) его проведения строго снизу вверх. Структуризация информационного анализа определяется четырьмя принципами.

**1. Принцип структурированности.** Означает, что анализ полного множества элементов иерархически структурируется по узлам анализа. Структуризация проводится в соответствии с некоторыми критериями равнозначности и связности элементов. Эти критерии определяют их распределение по уровням иерархии, включение в узлы анализа и связывание узлов по уровням.

В качестве критерия равнозначности может быть использована доля элемента в полном их множестве. В этом случае распределение элементов по уровням определяется величиной размаха долей и шагом его дискретизации, определяющим число уровней. Распределение элементов по одноуровневым узлам определяется посредством минимизации среднеквадратического отклонения долей элементов, отбираемых в конкретный узел анализа. При этом используются не доли элементов в полном их множестве, а доли элементов в отбираемой для узла группе. Естественно, доли элементов в некоторой их группе (подмножестве), вошедшей в узел анализа, будут существенно отличаться от долей этих же элементов, но в полном их множестве.

Для элементов, имеющих более одного показателя, использование доли в качестве критерия равнозначности может приводить к различным структурам информационного анализа, число которых определяется числом показателей. В этом аспекте с целью уменьшения структур анализа следует использовать комплексные критерии равнозначности по группам показателей. В отличие от критерия равнозначности, определяющего распределение элементов по уровням и узлам анализа, критерий связности обеспечивает установление связей между разноуровневыми узлами анализа. В качестве критерия связности может выступать территориальная близость предприятий, схожесть горно-геологических условий, марка углей, административная подчиненность и др.

**2. Принцип соразмерности.** Собственно говоря, этот принцип и определяет структуру информационного анализа, поскольку требует равных условий для сопоставления элементов. Нет смысла сравнивать и сопоставлять несоизмеримые вещи. Элементы, включаемые в узел анализа в соответствии с критерием равнозначности, должны быть соразмерными, например, сопоставимы (незначительно отличаться друг от друга) их узловые доли. В то же время анализ элементов в узле должен проводиться не по соотношению их узловых долей или комплексного соотношения их долей в случае множественности показателей, тем более, результирующего плана (добыча, доход, численность). Узловой анализ должен основываться на глубоком и объективном подходе, который в сопоставимых единицах могут отражать критерии (на примере угледобывающих предприятий):

- затрат - сопоставляются совокупные затраты предприятия на добычу тонны угля и объективно необходимые, обусловленные природным фактором,

- разведанности и подготовленности запасов в соотношении с уже отработанными запасами,

- оценки последствий добычи (экологический и социальный аспект),

- интеллектуальности - способность оптимально действовать в экстремальных ситуациях, вырабатывать и применять новые технологические решения и финансировать перспективные научные разработки,

- рыночной активности предприятия - реальные и потенциальные возможности экспансии внутреннего и внешнего рынков,

- финансовой стабильности предприятия.

**3. Принцип элементной ограниченности.** Накладывает ограничение на количество анализируемых в узле элементов. В соответствии с критерием элементной емкости единичной структуры это число выбирается из диапазона (*2-6*). Оптимальным является число три. Узел, содержащий 3 элемента, считается оптимальным, а узел, содержащий другое число элементов, - квазиоптимальным.

**4. Принцип последовательного выравнивания.** Регламентирует порядок проведения анализа, его направленность строго снизу вверх. Анализ начинается с узлов нижнего уровня, в которых определяются узловые лидеры, и последовательно переходит на вышестоящий уровень с выдвижением лидеров нижнего уровня в соответствующие узлы верхнего уровня, согласно установленных межуровневых связей. Следует отметить, что связность узлов разных уровней конкретизирует пути движения лидеров вверх и определяет одно направление для каждого лидера. Отсутствие связности между узлами двух уровней порождает многовариантность движения.

## СПИСОК ЛИТЕРАТУРЫ

1. *Логов А.Б., Поварницын В.И., Кочетков В.Н.* Моделирование состояния угольного комплекса Кузбасса на стадии реструктуризации. – Новосибирск: Издательство СО РАН, 1999. – 100 с.
2. *Острейковский В.А.* Информатика: Учеб. для вузов. – М.: Высш. шк., 1999.
3. Автоматизированные информационные технологии в экономике: Учебник /М.И. Семенов, И.Т. Трубилин, В.И. Лойко, Т.П. Барановская; Под ред. И.Т. Трубилина. – М.: Финансы и статистика, 2000.
4. *Шеннон К.* Работы по теории информации и кибернетике / Пер. с англ.; Под ред. Р.Л. Добрушина, О.Б. Лупанова. – М.: ИИЛ, 1963.

Автор статьи:

Преслер
Вильгельм Теобальдович
- докт.техн.наук, проф. каф. ИиАПС,
ведущий научный сотрудник Института
угля и углехимии СО РАН